DOCUMENT RESUME

ED 306 310 TM 013 258

TITLE Guide to Developing a Comprehensive Assessment Procedure for Your District.
INSTITUTION Regional Laboratory for Educational Improvement of the Northeast & Islands, Andover, MA.
SPONS AGENCY Department of Education, Washington, DC.
PUB DATE 87
CONTRACT 400-86-0005
NOTE 112p.
PUB TYPE Guides - Non-Classroom Use (055)

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS *Academic Achievement; Achievement Tests; Comparative Analysis; *Educational Assessment; Elementary Secondary Education; Evaluation Methods; *Evaluation Utilization; Guidelines; Performance; *Program Development; Program Evaluation; Public Schools; *School Districts; School Effectiveness; Standardized Tests; Test Results
IDENTIFIERS *Comprehensive Assessment Report; Evaluation Reports; New York

ABSTRACT

This guide describes the New York State assessment process as a framework for other districts aiming to develop an effective approach to performance assessment. In 1984-85 the New York State Board of Regents initiated the Comprehensive Assessment Report (CAR) as a process of reporting and using education results throughout the state. The conceptual approach of the guide conceives of standardized test data as a limited first step in assessing school effectiveness. In going beyond test data, one must look for convergence of test data and other indicators of school performance and one must separate the data into component parts to gain information about subgroups of students and schools. Although technical statistical considerations are not emphasized in this guide, it is essential that comparisons between schools, districts, or groups take into account demographic and other salient characteristics of the groups being examined. The introduction establishes the framework for this guide. The second chapter describes the CAR, its purposes and processes, and the data it includes. Chapter 3 discusses the limitations of the data and possible misuses. Chapter 4 generalizes from the New York experience to indicate how a district can prepare and present a comprehensive assessment report. Appendix 1 contains a model assessment report, and Appendix 2 focuses on basic measurement concepts. Appendix 3 presents a directory of school improvement resources, and Appendix 4 includes a 107-item bibliography of research on school effectiveness. Five data tables and 12 graphs are included. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

TM

ED306310

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it
☐ Minor changes have been made to improve reproduction quality

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
DAVID P. CRANDALL

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

GUIDE TO DEVELOPING A COMPREHENSIVE ASSESSMENT PROCEDURE FOR YOUR DISTRICT

1987
The Regional Laboratory

TM013258

BEST COPY AVAILABLE

This Guide was produced by The Regional Laboratory for Educational Improvement of the Northeast and Islands. It is based on work sponsored wholly or in part by the U.S. Department of Education under contract number 400-86-0005. The content of this publication does not necessarily reflect the views of the Department or any other agency of the U.S. government.

The Regional Laboratory for Educational Improvement
of the Northeast and Islands
290 South Main Street
Andover, MA 01810
508-470-0098

Acknowledgements

The Regional Laboratory acknowledges the advice, support, and expertise of many people and organizations that contributed to this Guide. Metis Associates, Inc., of New York City developed the Guide. Robert Crain, Ph.D., of Teachers College of Columbia University, contributed substantially to the development of the bibliography. Norm Fruchter of the Academy for Educational Development provided direction, advice, and support at every step of its development.

Paul W. Haley, at the Lab's Buffalo office, and Janet Angelis, Susan Loucks-Horsley and Leslie Hergert, also of The Regional Laboratory, critiqued drafts of the Guide and offered many useful suggestions pertaining to style, substance, and dissemination.

Richard Guttenberg of the New York City Board of Education encouraged its development, especially in the early stages, as did John Murphy and Bruce Crowder of the New York State Education Department.

The Guide was produced by the Regional Laboratory under the direction of Patrick Montesano.

Table of Contents

Preface i

Acknowledgements ii

1. Introduction 1

    Purpose of the Guide 2
    Conceptual Framework 2
    Organization of the Guide 3

2. New York State Comprehensive Assessment Report (CAR) 4

    Purpose of the CAR 4
    The CAR Process 4
    Data Included in the CAR 5

        Measures of Student Achievement 5

            Standardized Tests 5
            High School Graduation Results 10

        Demographic Information 12

            Racial/Ethnic Distribution and Limited English Proficiency 12
            Socioeconomic Indicators 13
            Pupil Mobility Indicator 13

        School Characteristics 14

            Enrollment, Attendance and Dropout Rates 14
            Average Class Size 14
            Pupil/Support Staff Ratio 15

3. Limitations of the CAR 16

4. Preparing and Presenting a Comprehensive Assessment Report 20

    Preparing the Assessment Report 20

        Step One: Verifying Data 20
        Step Two: Analyzing Trends 21
        Step Three: Supplementing the Data 26
        Step Four: Applying Appropriate Analytic Techniques 29

    Presenting the Assessment Report 30

Appendices.. 33

1. Model Comprehensive Assessment Report: Executive Summary............. 34

2. Basic Measurement Concepts 56

3. Directory of School Improvement Resources......... 65

4. Bibliography of School Effectiveness Research..... 82-89

Chapter 1

INTRODUCTION

As part of the nationwide emphasis on school improvement, schools and districts throughout the country are being asked to assess student achievement, to present data for public review, and to initiate district planning procedures to enhance school effectiveness.

In many instances, this attention to formal assessment has been a direct response to mandates emanating from state education agencies or state legislatures. Confronted with widespread concern about a decline in education standards and growing demands for accountability, numerous states throughout the country have initiated reforms that require and support public disclosure of student achievement data.

Presenting school/student achievement data is obviously not without risk even in the most politically stable school districts. The possibility that users will misinterpret the data are considerable, as are the possibilities that the data will correctly suggest problems that are politically costly to acknowledge and resolve.

Advocates say, however, that states and districts that open themselves to such scrutiny are taking a critical step toward unifying and consolidating community support for their schools and their school improvement efforts and strengthening the partnership essential for progress.

Expressing that hope, in 1984-85 the New York State Board of Regents initiated the Comprehensive Assessment Report (CAR). The CAR, which is part of the Regents Action Plan to Improve Elementary and Secondary Education in New York State, provides a process for reporting and using education results throughout the state.

Purpose of the Guide

This guide will describe in considerable detail the scope and limitations of the New York State assessment process and steps that can be taken to amplify CAR data and make them more useful for school improvement purposes. Our intention is both to create a practical guide of immediate value to schools and districts in New York State and to use the New York State experience to frame and highlight issues that could usefully be considered by districts outside the state aiming to develop an effective approach to performance assessment.

We have prepared the text with the needs of district-level administrators in mind. However, building administrators, teachers, school improvement team members, and parents may also benefit from the discussion. The presentation is largely nontechnical and the approach stresses questions that one should ask when examining performance data.

Conceptual Framework

Most guidebooks begin by providing a map of the territory the authors consider key to the phenomena they are discussing. Before launching into the substance of this guide, we would similarly like to provide a general framework to help readers understand why we suggest the things we do.

We see analysis of standardized test data as a limited first step in assessing school effectiveness - as an important but not sufficient indicator for determining whether schools are working effectively with students. To make effective use of standardized test data one must go beyond them. One must look for convergence between test data and other indicators of school performance, such as teachers' subjective reports of how students are doing. One must also carefully "disaggregate the data" or separate them into their component parts so that they provide information about subgroups of students and schools. When data are looked at only in summary form one often misses important differences that may exist by race, gender, students' poverty status, etc.

You will note when reading this guide that we very cautiously mention statistical procedures that districts can perform to determine whether obtained differences in student performance are significant mathematically, that is, unlikely to have occurred by chance. We don't stress these technical statistical considerations, however, for two reasons: first because a preoccupation with determining statistical significance may inhibit districts from a common sense examination of the data; second because when districts find similar trends suggested by several standardized tests and other indicators across time, this information will probably provide a meaningful basis for action.

What we do emphasize is that when schools, districts, and student groups are compared, those comparisons must take account of the demographic and other salient characteristics of the groups being examined. Without this check, it would be very difficult to identify schools with a high percentage of academically able students that are failing to maximize those children's potential. Similarly it would be difficult to identify schools with a high proportion of academically disadvantaged students that are making great progress with their youngsters.

Organization of the Guide

The guide is divided into four chapters. The first is this Introduction, which establishes the framework for the report. The second chapter describes the New York State Comprehensive Assessment Report (CAR), its purpose and processes, and the data included in the CAR. Chapter Three discusses the limitations of the data and possible misuses. Chapter Four generalizes from the New York State experience and details how a district might prepare and present a comprehensive assessment report. Such issues as data verification, supplementation, interpretation, and illustration are addressed in that chapter.

Appended to the guide are sections that present useful information and resources to support the comprehensive assessment and school improvement process. Appendix 1 contains a model assessment report. Appendix 2 focuses on understanding normative test data. Appendix 3 presents a directory of school improvement resources, including books, manuals, and organizational resources. And Appendix 4 is a bibliography of school effectiveness research.

Chapter 2

THE NEW YORK STATE COMPREHENSIVE ASSESSMENT REPORT (CAR)

Purpose of the CAR

The Comprehensive Assessment Report (CAR) is an essential part of the Regents Action Plan to Improve Elementary and Secondary Education in New York State.

The New York State Board of Regents expressed the hope that this opportunity for public review and discussion of student achievement would serve two ends:

- increase community involvement in education and
- assist district staff in planning for school improvement.

As stated in the Regents Action Plan:

> Public review and debate about student achievement has proven to be one of the most effective ways to develop the concern of parents and the public and to bring about action. From experience gained from the Department's Resource Allocation Plan (RAP), one of the most effective ways for drawing attention and involvement of communities and boards of education is to display trend information on pupil progress. The information is itself a powerful motivation to take action.

The CAR Process

Each year the New York State Education Department (SED) administers a number of standardized testing programs and obtains demographic and other data from each of its schools. These data are entered into the Basic Education Data System (BEDS). The CAR integrates much of this information. The data are aggregated at the school and district levels for three successive school years. Achievement, demographic, and other data are presented in a report to local districts. The districts are then required to review the information and report it to their Boards of Education and members of the education community.

In addition to serving as a mechanism for initiating school improvement planning at the local district level, the CAR data are used by the state to identify schools "most in need of assistance." Schools scoring in the bottom 10 percent of the distribution on one or another of the CAR criteria are identified for special attention. Such schools are required to develop a comprehensive school improvement plan (CSIP). The format of the plan is prescribed by the Commissioner. The plans themselves are developed by the schools with the assistance of State Education Department staff assigned to the particular school. Following completion, each plan must be approved by the local Board of Education and submitted to the State Education Department no later than April 30, with implementation to begin the following September.

Data Included in the CAR

The Comprehensive Assessment Report developed by the State Education Department provides three-year trend data that includes the following information:

- measures of student achievement;
- demographic information; and
- school characteristics.

The components of each data set will be elaborated fully in this chapter.

Measures of Student Achievement

The CAR presents two forms of student achievement data: standardized tests and high school graduation results.

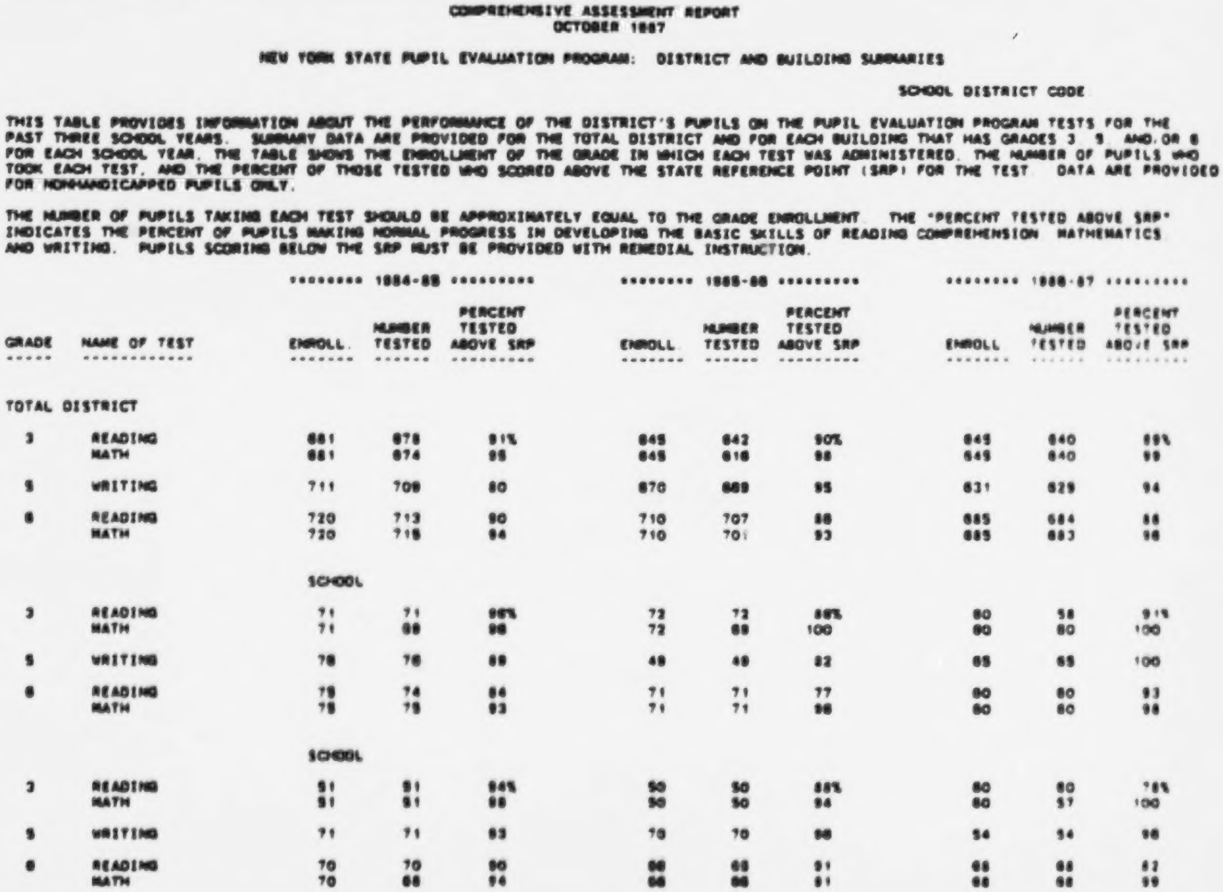
Standardized Tests. The New York State Testing Program is designed to assess student performance on a regular basis throughout the educational process. One of the primary purposes of this testing program is to assist schools in identifying students with learning difficulties.

The Comprehensive Assessment Report includes results from the Pupil Evaluation Program (PEP), the Program Evaluation Test (PET), the Preliminary Competency Tests (PCTs), the Regents Competency Tests (RCTs), and the Regents Examinations.

The PEP tests are administered at the elementary school level, the PCTs at the middle school level, and the RCTs and Regents Exams at the secondary level. Tests at the elementary and middle school level focus on reading, writing, and mathematics skills. The high school level exams measure academic achievement as well as basic skills competency.

Pupil Evaluation Program (PEP). Students in grades three and six take PEP tests in reading and math. In grade five, writing skills are assessed. The reading and math tests contain multiple-choice items that are revised every three years. The writing test is a two-part essay exam that is graded by local teachers in accordance with the state's holistic scoring guidelines.

The table that follows presents sample district and building summaries for the New York State Pupil Evaluation Program.



COMPREHENSIVE ASSESSMENT REPORT
OCTOBER 1987

NEW YORK STATE PUPIL EVALUATION PROGRAM: DISTRICT AND BUILDING SUMMARIES

SCHOOL DISTRICT CODE

THIS TABLE PROVIDES INFORMATION ABOUT THE PERFORMANCE OF THE DISTRICT'S PUPILS ON THE PUPIL EVALUATION PROGRAM TESTS FOR THE PAST THREE SCHOOL YEARS. SUMMARY DATA ARE PROVIDED FOR THE TOTAL DISTRICT AND FOR EACH BUILDING THAT HAS GRADES 3, 5, AND/OR 6 FOR EACH SCHOOL YEAR. THE TABLE SHOWS THE ENROLLMENT OF THE GRADE IN WHICH EACH TEST WAS ADMINISTERED, THE NUMBER OF PUPILS WHO TOOK EACH TEST, AND THE PERCENT OF THOSE TESTED WHO SCORED ABOVE THE STATE REFERENCE POINT (SRP) FOR THE TEST. DATA ARE PROVIDED FOR NONHANDICAPPED PUPILS ONLY.

THE NUMBER OF PUPILS TAKING EACH TEST SHOULD BE APPROXIMATELY EQUAL TO THE GRADE ENROLLMENT. THE "PERCENT TESTED ABOVE SRP" INDICATES THE PERCENT OF PUPILS MAKING NORMAL PROGRESS IN DEVELOPING THE BASIC SKILLS OF READING COMPREHENSION, MATHEMATICS AND WRITING. PUPILS SCORING BELOW THE SRP MUST BE PROVIDED WITH REMEDIAL INSTRUCTION.

		******** 1984-85 *********			******** 1985-86 *********			******** 1986-87 *********		
GRADE	NAME OF TEST	ENROLL.	NUMBER TESTED	PERCENT TESTED ABOVE SRP	ENROLL.	NUMBER TESTED	PERCENT TESTED ABOVE SRP	ENROLL.	NUMBER TESTED	PERCENT TESTED ABOVE SRP
TOTAL DISTRICT										
3	READING	881	878	91%	845	842	90%	845	840	89%
	MATH	881	874	95	845	816	98	845	840	99
5	WRITING	711	709	80	670	669	95	631	629	94
6	READING	720	713	90	710	707	88	685	684	88
	MATH	720	715	94	710	701	93	685	683	98
		SCHOOL								
3	READING	71	71	96%	72	72	88%	60	58	91%
	MATH	71	68	98	72	69	100	60	60	100
5	WRITING	78	76	88	49	49	82	65	65	100
6	READING	75	74	84	71	71	77	60	60	93
	MATH	75	75	93	71	71	98	60	60	98
		SCHOOL								
3	READING	51	51	86%	50	50	88%	60	60	78%
	MATH	51	51	98	50	50	94	60	57	100
5	WRITING	71	71	83	70	70	98	54	54	98
6	READING	70	70	90	68	68	91	66	66	82
	MATH	70	68	94	68	68	81	66	66	99

As the table indicates, for each grade and test the CAR presents the following information from the PEP tests:

- **Enrollment:** The number of non-handicapped students in the specified grade at the time of the test.

- **Number Tested:** The number of students who actually took the test. Absences as well as exemptions due to limited

English proficiency account for the difference between "enrollment" and "number tested."

- **Percent Tested Above SRP:** The percent of students making normal progress in developing the basic skills. The State Reference Point (SRP) is the state's criterion of satisfactory performance. The SRP is based on the SED's calculation of test score averages across the state during the first year the exam is administered. A score below the SRP is an indication that a student requires special help or remediation in the skill area. Students scoring below the SRP must be provided with remediation.

Program Evaluation Tests (PETs). The PET in social studies is administered annually to pupils in grade 6 beginning in the 1986-1987 school year. The test is designed to measure the effectiveness of the elementary school social studies programs of public and nonpublic schools within the state. The test yields data that can be used in the planning, management, and evaluation of programs at both state and local levels. The test contains three components, two of which are required and are reported on the CAR. The two required components are the objective test, consisting of fifty questions, and the writing sample, involving one essay. The third component, the participation project, is optional. The participation project offers schools an opportunity to assess how well pupils apply content, concept, and skill learning in a group problem solving situation.

The CAR contains data for the PET for the total district as well as for each school that administered the test. Data are presented for nonhandicapped pupils only. Three columns of data are provided. The first column indicates the number of nonhandicapped pupils tested. The second column indicates the mean score for the pupils tested. The third column indicates the percentile rank associated with the mean score. The percentile ranks are based on a frequency distribution of mean scores for all public and nonpublic schools in New York State administering the grade 6 social studies test. Thus, when percentile ranks are reported for a public school district, the data represent the percentile ranks of the "average school" in that district.

Preliminary Competency Tests (PCTs). These tests, given to students in either grade 8 or 9, are also designed to assess basic skills in reading, mathematics, and writing. Like the PEP tests, their primary purpose is the identification of pupils who need special help if they are to develop a basic level of competence in these critical skill areas.

PCT results are presented in the CAR in the manner described for the PEP, however the PCT results include the Percent of Enrollment Above SRP, rather than the Percent Tested Above SRP. This statistic shows the percentage of students in the entire grade who scored above the state reference point. This percent includes those pupils who scored above the SRP on the Preliminary Competency Test as well as those who were exempted from the testing program because their competency had been established on other tests.

Regents Competency Tests (RCTs). In order to receive a local high school diploma, pupils must demonstrate competency in reading, writing and mathematics. The Regents Competency Tests (RCTs) provide one means for demonstrating the required level of proficiency. The first Regents Competency Test in mathematics was administered in June of 1979 and the first Regents Competency Tests in reading and writing were administered in January 1980. These tests established new standards of achievement in basic skills for pupils receiving high school diplomas beginning with the class of 1981. The Regents Action Plan specifies mandates for competency tests in new curriculum areas. As of 1991, students will also be required to demonstrate competence in occupational education, science, global studies, and American government.

In most schools throughout New York State, the RCTs are taken only by those pupils who do not take Regents Examinations. Students who are enrolled in Regents level or advanced courses will normally choose to demonstrate competency by successful completion of the Regents examinations in English and Math.

Like the PEP and PCT programs, the RCTs consist of three tests in the same basic skills areas: reading, writing, and mathematics. The RCTs in mathematics are normally taken in the ninth grade but may be taken any time from grades nine to twelve. The RCTs in reading and writing may be taken only by pupils in grades eleven or twelve. Students who fail the RCT may repeat the test at a later date. In the interim, the district is required to provide appropriate remedial instruction.

Like the tests in the PEP and PCT groups, the RCT reading and math tests have a multiple-choice format. The writing test requires students to complete three writing tasks, each of which is assessed by one of three raters. Those writing tests that receive passing grades are then forwarded to SED for a second review and final pass/fail decision.

As with the PEP, PET and PCT reports, the reports for the Regents Competency Tests also provide information for the entire district and for individual buildings for each of three years. The results reported are the total number of students tested (Number Tested) and the percent of students who passed (Percent Passed), that is, scored above the state reference point. The exams are given three times a year, in January, June, and August, but only January and June scores are reported.

Regents Examinations. Regents Examinations are achievement tests based on the courses of study recommended by the State Education Department for grades 9-12. Tests are currently offered in English, mathematics, foreign languages, sciences, social studies, and business.

Many high schools have rigorous admissions requirements for Regents courses, and the exams are taken primarily by college-bound students. The present Regents Examination Program began in 1878, and the tests have traditionally been used to gauge teaching effectiveness and learning success and to determine eligibility for the special Regents honors diploma.

Students may repeat any of the exams if they fail. Each exam is developed by the SED, working closely with classroom teachers. Most Regents Examinations combine multiple choice with essay questions, but the tests in chemistry, earth science, and physics are entirely multiple choice. All Regents Examinations are scored by local classroom teachers; random samples from each discipline are sent to the SED for a second review.

The CAR presents Regents Examination results for each individual test. As with the other statewide tests, information is presented for the total district and for individual school buildings. The report also indicates the number of students tested (Number Tested) and the percent passing of the total number tested (Percent Passing of No. Tested). In addition, the Regents Examination report also includes an additional column - Percent of Enrollment Tested. This number represents the relationship between the number of pupils taking the exam and the average enrollment for that particular grade. For example, assume that 200 tenth grade students completed the Social Studies test in June. In June, the total enrollment in that class was 400. Therefore, the Percent of Enrollment Tested would be 50%, i.e. 200 students out of a total of 400 student were tested at that time.

The following table presents an overview of the tests included in the CAR:

Overview of Tests Included in the CAR

Test	Skill Area	Grade Level
Pupil Evaluation Program	Basic Skills Testing:	3,5, and 6
	Reading	3 and 6
	Mathematics	3 and 6
	Writing	5
Program Evaluation Test	Social Studies	6
Preliminary Competency Test	Basic Skills Testing	8 or 9
	Reading	8 or 9
	Writing	8 or 9
Regents Competency Test	Basic Skills Testing	9 - 12
	Reading	11 or 12
	Writing	11 or 12
	Mathematics	9 - 12
Regents Examinations	All Academic Subjects and Business	9 - 12

High School Graduation Results. This section of the CAR indicates how many handicapped and non-handicapped students in the twelfth grade were candidates for graduation, and how many of these students actually completed the requirements entitling them to receive a high school diploma or certificate. The table that follows illustrates district and building summaries of high school graduation results for a sample district.

For each district and each high school within that district, the CAR lists the following for both handicapped and non-handicapped pupils for each of the three years: number of diploma candidates; number of candidates who received a local high school diploma; number of candidates who received a Regents diploma in addition to a local high school diploma; number of candidates who did not receive a local diploma; and number of pupils with handicapping conditions who received an IEP diploma or certificate.

Diploma candidates. Diploma candidates are pupils who, one month before graduation, should theoretically be able to complete all required testing and courses. For 1984-85, this figure includes only pupils who were diploma candidates on June 1. The 1985-86 figure includes those who were candidates any time during the year, including the summer.

COMPREHENSIVE ASSESSMENT REPORT
OCTOBER 1987

HIGH SCHOOL GRADUATION RESULTS: DISTRICT AND BUILDING SUMMARIES

SCHOOL DISTRICT CODE

THIS TABLE PROVIDES INFORMATION ABOUT THE DISTRICT'S PUPILS WHO WERE CANDIDATES FOR GRADUATION DURING EACH OF THE PAST THREE SCHOOL YEARS. FOR THE 1984-85 SCHOOL YEAR, SUMMARY DATA ARE PROVIDED ONLY FOR THE TIME PERIOD BETWEEN SEPTEMBER 1 AND JUNE 30; DATA ARE NOT INCLUDED FOR PUPILS WHO MET THE REQUIREMENTS FOR GRADUATION DURING JULY AND AUGUST. FOR THE 1985-86 AND 1986-87 SCHOOL YEARS, SUMMARY DATA ARE PROVIDED FOR THE TIME PERIOD BETWEEN SEPTEMBER 1 AND AUGUST 31. THE DATA FOR THE 1985-86 AND 1986-87 SCHOOL YEARS INCLUDE ALL PUPILS WHO MET THE REQUIREMENTS FOR GRADUATION PRIOR TO AUGUST 31, EVEN THOUGH SOME PUPILS MAY NOT HAVE RECEIVED THEIR DIPLOMAS UNTIL AFTER AUGUST 31.

SUMMARY DATA ARE PROVIDED FOR THE TOTAL DISTRICT AND FOR EACH BUILDING THAT CONTAINED GRADE 12 PUPILS. THE TABLE SHOWS THE NUMBER OF PUPILS WHO WERE DIPLOMA CANDIDATES AS OF JUNE 1 FOR THE 1984-85 SCHOOL YEAR, AND AT ANY TIME BETWEEN SEPTEMBER 1 AND AUGUST 31 FOR 1985-86 AND 1986-87. THE TABLE ALSO SHOWS THE NUMBER OF PUPILS WHO RECEIVED LOCAL HIGH SCHOOL DIPLOMAS DURING EACH SCHOOL YEAR, THE NUMBER WHO RECEIVED REGENTS DIPLOMAS, AND THE NUMBER WHO FAILED TO GRADUATE BY THE END OF THE SCHOOL YEAR. SEPARATE DATA FOR HANDICAPPED AND NONHANDICAPPED PUPILS ARE NOT AVAILABLE FOR THE 1984-85 SCHOOL YEAR.

FINALLY, THE TABLE SHOWS THE NUMBER OF HIGH SCHOOL INDIVIDUALIZED EDUCATION PROGRAM (IEP) DIPLOMAS AND CERTIFICATES AWARDED TO PUPILS WITH HANDICAPPING CONDITIONS WHO WERE UNABLE TO COMPLETE THE REQUIREMENTS FOR A REGULAR LOCAL HIGH SCHOOL DIPLOMA. ONLY HANDICAPPED PUPILS WHO CANNOT COMPLETE THE REQUIREMENTS FOR A REGULAR DIPLOMA ARE ELIGIBLE FOR A HIGH SCHOOL IEP DIPLOMA OR A CERTIFICATE. DATA FOR THE NUMBER OF PUPILS RECEIVING IEP DIPLOMAS ARE NOT AVAILABLE FOR THE 1984-85 SCHOOL YEAR.

IN THE TYPICAL HIGH SCHOOL, ABOUT 92 PERCENT OF THE CANDIDATES RECEIVE A HIGH SCHOOL DIPLOMA AND ABOUT 43 PERCENT OF THE GRADUATES RECEIVE A REGENTS DIPLOMA. THE PRIMARY REASON FOR MISSING GRADUATION IS FAILURE TO SATISFY A LOCAL COURSE REQUIREMENT

	1984-85			1985-86			1986-87		
	NONHANDI-CAPPED	HANDI-CAPPED	TOTAL	NONHANDI-CAPPED	HANDI-CAPPED	TOTAL	NONHANDI-CAPPED	HANDI-CAPPED	TOTAL
TOTAL DISTRICT									
1. NO. OF DIPLOMA CANDIDATES DURING SCHOOL YEAR	NA	NA	995	917	36	953	961	15	976
2. NO. OF CANDIDATES WHO RECEIVED LOCAL HIGH SCHOOL DIPLOMAS	NA	NA	912	887	34	921	908	14	922
3. NO. OF CANDIDATES WHO RECEIVED REGENTS DIPLOMAS	NA	NA	389	341		341	414		414
4 NO. OF CANDIDATES WHO, BY THE END OF THE SCHOOL YEAR, FAILED TO SATISFY:									
- LOCAL COURSE REQUIREMENTS ONLY	NA	NA	76	2		2	49	1	50
- COMPETENCY TESTING REQUIREMENTS ONLY	NA	NA	5	2	2	4	3		3
- BOTH OF THE ABOVE REQUIREMENTS	NA	NA	2	26		26	1		1
5. NO. OF HANDICAPPED PUPILS WHO RECEIVED IEP DIPLOMAS		NA			4			6	
6. NO. OF HANDICAPPED PUPILS WHO RECEIVED CERTIFICATES		6							

	1984-85			1985-86			1986-87		
	NONHANDI-CAPPED	HANDI-CAPPED	TOTAL	NONHANDI-CAPPED	HANDI-CAPPED	TOTAL	NONHANDI-CAPPED	HANDI-CAPPED	TOTAL
S H S									
1. NO. OF DIPLOMA CANDIDATES DURING SCHOOL YEAR	NA	NA	545	497	18	515	439	8	447
2. NO. OF CANDIDATES WHO RECEIVED LOCAL HIGH SCHOOL DIPLOMAS	NA	NA	505	497	18	515	410	7	417
3. NO. OF CANDIDATES WHO RECEIVED REGENTS DIPLOMAS	NA	NA	236	175		175	182		182
4. NO. OF CANDIDATES WHO, BY THE END OF THE SCHOOL YEAR, FAILED TO SATISFY:									
- LOCAL COURSE REQUIREMENTS ONLY	NA	NA	34				29	1	30
- COMPETENCY TESTING REQUIREMENTS ONLY	NA	NA	5						
- BOTH OF THE ABOVE REQUIREMENTS	NA	NA	1						
5. NO. OF HANDICAPPED PUPILS WHO RECEIVED IEP DIPLOMAS		NA			3			3	
6. NO. OF HANDICAPPED PUPILS WHO RECEIVED CERTIFICATES		2							

Number of candidates who received a local high school diploma. The local diploma is the basic credential signifying completion of a high school program. The Regents diploma, or Regents endorsement of the local diploma, is an honor accorded to those pupils who meet additional examination and course requirements. The number in this section of the CAR represents all students who completed requirements for graduation, including those who received a Regents (honors) diploma.

Number of candidates who received a Regents diploma in addition to a local high school diploma. Normally, the requirements for a Regents diploma are more stringent than those for the local high school diploma. Where a student has completed the additional requirements necessary for the Regents diploma, this diploma will be awarded in addition to the local high school diploma.

Number of candidates who did not receive a local diploma. This identifies the number of students who were unable to complete requirements for the local diploma because they failed to satisfy: course requirements, competency test requirements, or both course requirements and competency test requirements.

Number of pupils with handicapping conditions who received an IEP diploma or certificate. Schools are required to develop an individualized education program (IEP) for each student who is classified as handicapped. This IEP establishes diploma requirements that are tailored to the specific needs and abilities of the student. In many cases, the handicap precludes completion of all of the requirements for either the Regents or local diploma. For these students, other requirements are established and the successful completion entitles them to receive a diploma or certificate.

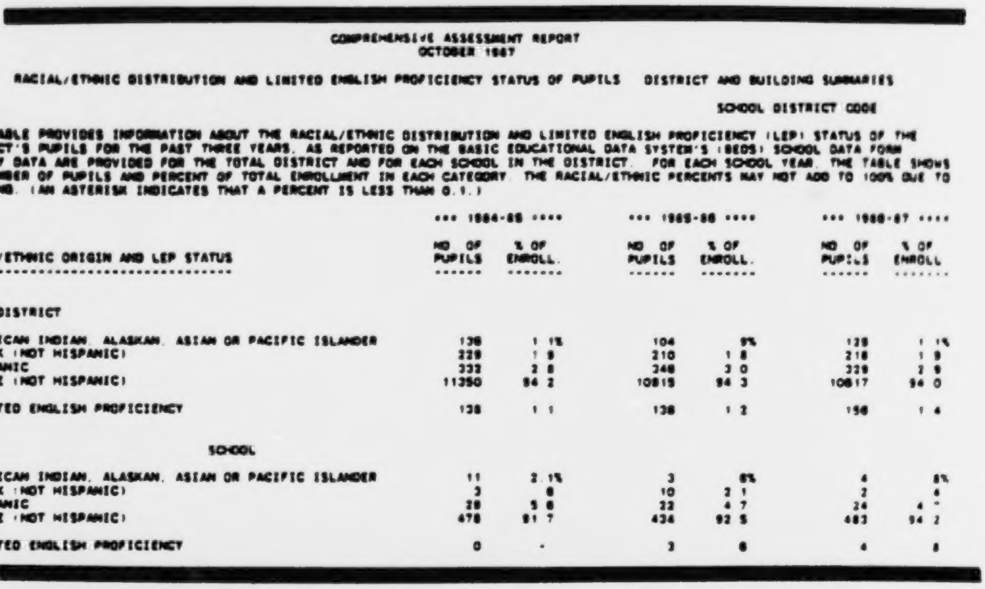
Demographic Information

The CAR presents information about the following demographic characteristics of the student population.

Racial/Ethnic Distribution and Limited English Proficiency. This section of the CAR report provides information about the racial/ethnic distribution and limited English proficiency (LEP) status of pupils in the district and in each school.

- Racial/ethnic distribution data include the number and percent of pupils belonging to each of four racial/ethnic groups: American Indian, Alaskan, Asian or Pacific Islanders; Black (non-Hispanic); Hispanic; and White (non-Hispanic).

- **Limited English Proficiency** (LEP) refers to the number and percent of students in the district or school who are not fluent in the English language. According to the New York State Commissioner of Education, students with limited English proficiency are students who speak a language other than English and score below the 23rd percentile on an English language assessment instrument approved by the Commissioner. The CAR provides racial/ethnic distribution and limited English proficiency data for the past three years. These data are illustrated in the table that follows.



COMPREHENSIVE ASSESSMENT REPORT
OCTOBER 1987

RACIAL/ETHNIC DISTRIBUTION AND LIMITED ENGLISH PROFICIENCY STATUS OF PUPILS DISTRICT AND BUILDING SUMMARIES

SCHOOL DISTRICT CODE

THIS TABLE PROVIDES INFORMATION ABOUT THE RACIAL/ETHNIC DISTRIBUTION AND LIMITED ENGLISH PROFICIENCY (LEP) STATUS OF THE DISTRICT'S PUPILS FOR THE PAST THREE YEARS, AS REPORTED ON THE BASIC EDUCATIONAL DATA SYSTEM'S (BEDS) SCHOOL DATA FORM. SUMMARY DATA ARE PROVIDED FOR THE TOTAL DISTRICT AND FOR EACH SCHOOL IN THE DISTRICT. FOR EACH SCHOOL YEAR, THE TABLE SHOWS THE NUMBER OF PUPILS AND PERCENT OF TOTAL ENROLLMENT IN EACH CATEGORY. THE RACIAL/ETHNIC PERCENTS MAY NOT ADD TO 100% DUE TO ROUNDING. (AN ASTERISK INDICATES THAT A PERCENT IS LESS THAN 0.1.)

RACIAL/ETHNIC ORIGIN AND LEP STATUS	*** 1984-85 *** NO. OF PUPILS	% OF ENROLL.	*** 1985-86 *** NO. OF PUPILS	% OF ENROLL.	*** 1986-87 *** NO. OF PUPILS	% OF ENROLL.
TOTAL DISTRICT						
AMERICAN INDIAN, ALASKAN, ASIAN OR PACIFIC ISLANDER	130	1.1%	104	.9%	128	1.1%
BLACK (NOT HISPANIC)	229	1.9	210	1.8	218	1.9
HISPANIC	332	2.8	348	3.0	329	2.9
WHITE (NOT HISPANIC)	11350	94.2	10815	94.3	10817	94.0
LIMITED ENGLISH PROFICIENCY	138	1.1	138	1.2	158	1.4
SCHOOL						
AMERICAN INDIAN, ALASKAN, ASIAN OR PACIFIC ISLANDER	11	2.1%	3	.6%	4	.8%
BLACK (NOT HISPANIC)	3	.6	10	2.1	2	.4
HISPANIC	28	5.8	22	4.7	24	4.7
WHITE (NOT HISPANIC)	478	91.7	434	92.5	483	94.2
LIMITED ENGLISH PROFICIENCY	0	-	3	.6	4	.8

Socioeconomic Indicators. Two socioeconomic indicators are provided in the report:

- **Socioeconomic Indicator (Census)** refers to the percentage of children, aged 5-17, whose families are below the poverty level as determined by the 1980 Federal Census. This figure is reported for the total district only and, since it is determined using information obtained in the 1980 Federal Census, the percentage will remain the same until the completion of the next census.

- **Socioeconomic Indicator (BEDS)** shows the percentage of pupils in each school who are members of families whose primary means of support is a public welfare program. The numbers are based on data submitted to SED by the school over the past three years and are presented as a range, e.g., 1-10%, 11-20%, 21-30%, 31-40%.

Pupil Mobility Indicator. The pupil mobility indicator shows the percent of pupils in the highest grade of each school who were enrolled in that same school during the previous year. For example, in an elementary school with six grades, the pupil mobility indicator shows the percent of sixth grade students who were enrolled in the school in the fifth grade. Information is provided for each school building in the district and data are reported as a range, e.g., 1-10%, 11-20%, 21-30%, 31-40%.

School Characteristics

In this category, the CAR incorporates the following descriptive information: enrollment, attendance, and dropout rates; average class size; and pupil/support staff ratio.

Enrollment, Attendance, and Dropout Rates. The CAR contains information on enrollment and attendance by grade and dropout rates for grades 9-12 for the total district and for each school, for the past three years.

- Enrollment rates are based on fall reports and may differ from enrollment numbers reported for the PEPs and PCTs, which are given later in the school year.

- Attendance rates present the total number of days that pupils actually attended school as a percentage of the total number of days that school was in session. Since annual attendance cannot be calculated until the end of the school year, figures are not included for the current school year.

- Dropout rates are reported for the total district and for each school that contains any of grades 9-12. The report shows the number of students in grades 9-12 who left school and their percentage of the total enrollment in those grades. It is an annual dropout rate. The term "dropout," which is defined differently in other states and localities, when used by the New York State Education Department refers to any student who leaves school prior to graduation for any reason other than death and does not enter another school, alternative high school, equivalency program, or other diploma program. The number of students in the district who leave school but enter an alternative high school equivalency or other diploma program is also shown in the CAR report. This number is represented as a percentage of the total 9-12 enrollment.

Average Class Size. Average class size data are presented for elementary grades at the kindergarten and at the common branches (1-6) levels. For the middle school grades (7 and 8), average class sizes are provided for grade 8 classes in English, mathematics, science, and social studies. At the high school level, average class size information is presented for grade 10 classes in English, mathematics, science, and social studies. These grades and classes were selected to be representative of those with the largest typical enrollments.

Pupil/Support Staff Ratio. The pupil/support staff ratio refers to the number of support staff personnel available to provide services to students in the total district. The ratio expresses this relationship by showing the number of students to be served by each member of the support staff. This number is obtained by dividing the total district enrollment in grades K-12 by the total number of full-time staff (FTEs) classified as Support Staff. "Support staff" refers to guidance counselors, nurses, psychologists, psychiatrists, dental hygienists, librarians, social workers, and attendance teachers. (New York City, Buffalo, and Rochester do not include nurses and dental hygienists in this category.)

Chapter 3

LIMITATIONS OF THE COMPREHENSIVE ASSESSMENT REPORT (CAR)

New York State's efforts to provide districts with a broad range of information constitutes an important service; however, its usefulness in determining the educational effectiveness of schools should not be overestimated.

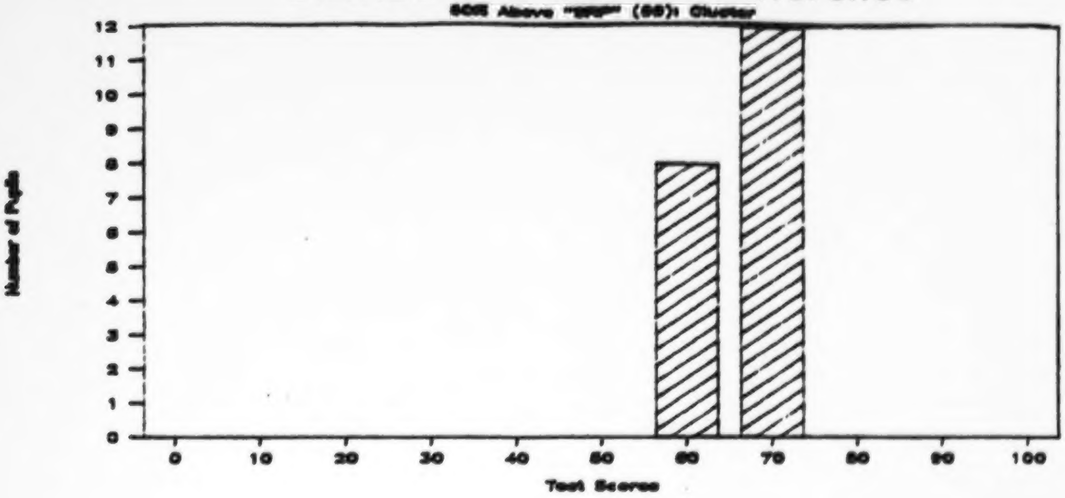
In this chapter we discuss some of the limitations of the CAR. The goal of this discussion is to provide a context in which districts can understand the importance of supplementing CAR data.

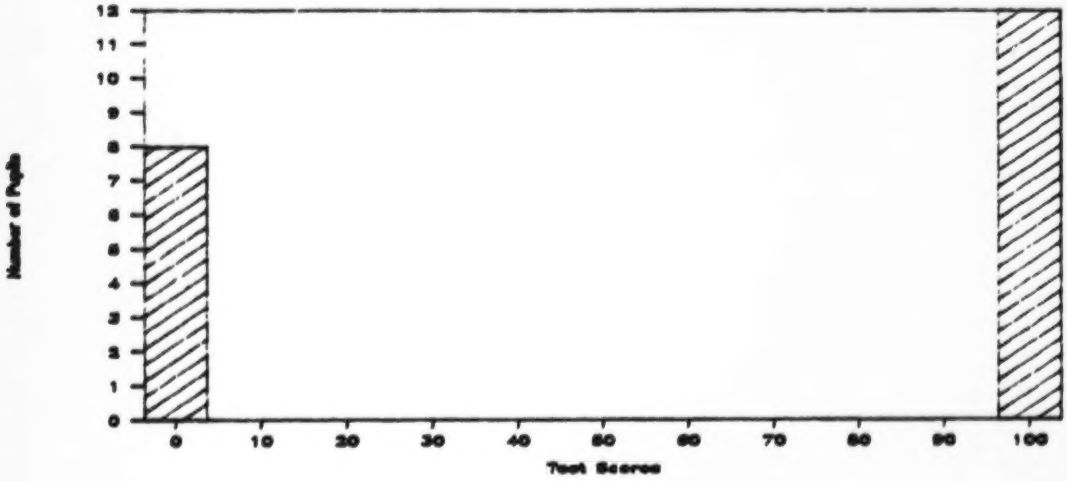
The New York State Regents intended the CAR to provide information that would encourage local school districts to focus on improving the educational process. It would be the initial stimulus for districtwide assessment and planning. Problems appear to arise when the CAR is viewed, not as the first step, but as the final step in this process. Some specific limitations are enumerated below.

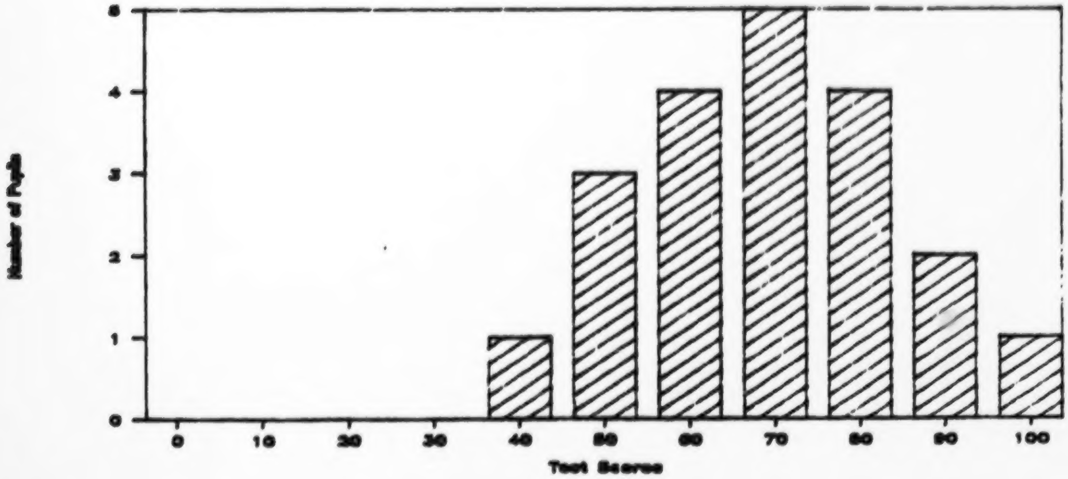
First, the data contained in the CAR present only a narrow view of any district's performance. The report focuses only on specific instructional objectives and outcomes and describes outcomes using gross and limited measures of central tendency, for example, the percent of students above the statewide reference point.

Assessments restricted to such statistics as "percent above SRP" can be very misleading. The graphs on the following page illustrate this point. They show a distribution of scores with 60 percent of the students achieving above the SRP. In this example the SRP is equal to a score of 69, and there were twenty students tested. In the top figure the test scores are clustered near the statewide reference point. In the middle figure the scores are extreme -- either zero or 100. The bottom figure shows the data to be fairly normally distributed between 40 and 100. Clearly the pedagogical response to each of these distributions would vary. Yet the "60 percent above the SRP" characterization obfuscates the difference.

Distributions Make a Difference


60% Above "SRP" (65): Cluster


60% Above "SRP": Extremes Distribution


60% Above "SRP": "Normal" Distribution

The second serious difficulty is that the CAR consistently presents data for different groups of students over three consecutive school years (e.g. third graders in 1983-84, third graders in 1984-85, and third graders in 1985-86). Reasonable assessments require, at a minimum, data for contiguous grades (i.e. third grade in year one, fourth grade in year two, fifth grade in year three). Ideally, the same students should be followed over time through longitudinal analyses in order to develop and test hypotheses about student outcomes related to multi-year "school effects." Examples of contiguous grade and longitudinal cohort analyses are contained in Appendix 1.

A third difficulty is that CAR data are not uniformly derived. Although the SED has established guidelines to facilitate uniform data collection and test administration procedures for all districts of the state, differing local policies and practices reduce the comparability of the obtained data. Policies vary with respect to several matters: which students get included or excluded from components of the testing program; how students get classified as LEP or handicapped; which students take practice exams; the extent to which school-provided data are validated; etc. Consider some specific illustrations.

- Districts differ considerably in the numbers of limited English proficient students who are included in the test taking population. The test manual requires that students receiving instruction in schools in the United States for two years take exams in English. Districts that have substantial numbers of immigrants are at a marked disadvantage as the language skills of new speakers of English, despite instruction, will likely lag behind native speakers' skills. To demonstrate this point further, consider New York City's annual ranking of schools by students' performance on reading tests. Schools are ranked on the percent of students achieving above grade level. For this calculation, however, students who have been excused from the test (including those excused for limited proficiency in English) are counted as below grade level.

- Handicapped students' test scores are reported separately from non-handicapped students' test scores in the CAR. While the term "handicapped" clearly refers only to those students who have been classified as such through the state-approved Committee on Special Education, the practice of testing and classifying students does, in fact, vary considerably by district. Some districts more effectively identify learning disabled youngsters than others. Consequently there are fewer lower ability youngsters' test scores factored into the general education data. Some districts routinely wait until after the third grade PEP test to classify students as learning disabled. Consequently test scores appear to improve in the later grades.

- Variation in the policy of exempting students from the examinations also affects test scores. In some districts students do not take Preliminary Competency Tests (PCTs). As a consequence scores on the actual RCTs may not be as high. Similarly many students take the RCTs in January as well as June, which tends to result in higher June test scores. Additionally, in school districts where a high proportion of the students take Regents Examinations rather than competency tests, scores on RCTs will be lower than expected because the exam is eschewed by more able students.

Fourth and finally, record-keeping practices vary within the state. Procedures for collecting data are better in some schools and districts than others. As a result, it is difficult to make valid comparisons between districts. Records are less likely to be reliable in areas with high pupil mobility.

Many of the limitations described above are not specific to the CAR but are general problems that educators must confront when utilizing standardized test data. In the chapter that follows we discuss steps analysts can take to overcome these problems.

Chapter 4

PREPARING AND PRESENTING A COMPREHENSIVE ASSESSMENT REPORT

The New York State Comprehensive Assessment Report is, as we have suggested, a useful but not sufficient tool for assessing school and student performance. This chapter explores steps that districts can take to improve upon the information made available by their state education departments in the preparation of a final local report.

Three general principles have guided us in developing these recommendations. First, we believe districts would be advised to examine carefully the data provided by the state to see if the data are compatible with information they already have about themselves.

Second, we believe districts would benefit from disaggregating or separating the state's data into their component parts so that the data yield information about appropriate subgroups of students, schools, subject areas, etc. The New York State CAR reports information in a way that makes comparisons between schools, between school years (e.g. 1986 and 1987), between academic grades (e.g. third grade performance and sixth grade performance), and between academic disciplines (reading, math, writing) feasible.

Third, we recommend that when districts make comparisons among schools, among grade levels, and among academic disciplines, etc., that they attempt to control for pre-existing differences in the groups being compared. At a minimum districts should attempt to find out if the students within the groups being compared were roughly comparable before the treatment, i.e., before enrollment in the grade level, subject area, or school in question.

The material that follows will elaborate these principles. It is the goal of this guidebook to help districts work with the data provided by their state education departments and move toward a richer and fuller local assessment report.

Preparing the Assessment Report

Step One: Verifying Data

In New York State, districts are required to submit demographic and institutional data to the state education department. The state compiles this information, and it is ultimately returned to schools. In the course of these transactions, errors are inevitably made. Data are occasionally lost; data are occasionally inaccurately or incompletely recorded. There is a clear need to check the state's information for accuracy and completeness. Missing data must be entered. Data that are inconsistent with other sources of information on the same subject should be scrutinized to determine if they are erroneous or if alternative methods of calculating outcomes simply revealed a different picture of the same phenomena.

The first step in the process of verification is to gather together the material the district originally sent to the state in response to the state's request for information and check the transcription of that material for accuracy. Are enrollment figures, class size data, racial breakdowns, etc., correct?

Additionally, if your school district has a local standardized testing program, it may wish to compile the results from those exams in a form parallel to the state's. If the state reports district test data in aggregate form and grade by grade, attempt to organize your own data in that fashion. Compare the state's information carefully with your own, checking for inconsistencies as a way of identifying possible errors.

To illustrate the subtlety of comparisons across tests, imagine the state reports 70% of your third graders at or above criterion on the Degrees of Reading Power (DRP) Test and that your own data, let us say from the California Achievement Test (CAT), show 50% of third graders at or above the national median. One could reasonably think these data were incompatible. In the process of attempting to verify the scores, you might find out from the state education department that the percent passing the statewide reference point on the DRP was generally higher than the percent at or above the national median on the California Achievement Test. This finding would tend to support the notion that the data were accurate and compatible. On the other hand, you might find that the percent passing the DRP criterion was generally lower than the percent above the national norm on the CAT. Such a discrepancy could signal an error, although there are certainly other possibilities. One would persist, in the latter case, until certain that the data were correctly reported.

Step Two: Analyzing Trends

After the possibility of error has been ruled out, the analyst's next task is to highlight and explain inconsistencies in the data. If, for example, the state's assessment suggested there were considerably fewer children reading below level than the district's assessment suggested, that information must be reconciled. It may be that one test had a lower criterion or "passing" grade than the other. It may be that different skills were measured by the separate tests or that different students were tested. Whatever the case, it is the analyst's job to point out inconsistencies and attempt to reconcile all available data.

When inconsistencies are satisfactorily reconciled, the process of disaggregating the data or subdividing them into their component parts and identifying trends begins. This is a painstaking task that requires considerable care and time. In the sections that follow, we will describe the process using data provided by the New York State Comprehensive Assessment Report.

The CAR presents data in a manner that facilitates time-ordered, within-district comparisons between schools, academic levels (e.g. third grade and fourth grade), and academic content areas (e.g. reading and math), as well as comparisons between an individual district and the state. Approaches toward these will be discussed.

District Level. The threshold analytic question for the district is how is it doing in comparison to the rest of the state. In New York State, the literal answer to that question is set by the percent of district students who perform above the Statewide Reference Point (SRP) in each skill area. The CAR makes it possible for each district to compare itself to the statewide average. A more sensitive comparison, however, would consider how the district is doing relative to districts with comparable student groups and comparable expenditures and in comparison to its own past performance. This information is also available.

After understanding how it is doing in comparison to the state, the district will be interested in determining if there is variation in its schools' effectiveness. Are students more likely to be successful in one district school than another? Again, it is not difficult to give a literal answer to that question. One can simply examine and compare the percent of students above the SRP in each school. Figure 1 on the following page illustrates this approach to school-to-school-and-district comparisons. In that figure third grade PEP reading performance is shown for a hypothetical district and for each of its six elementary schools.

Such a graph is easily generated from information available in the CAR. In school districts where the student population is generally homogeneous and expenditures per school are comparable, it is likely to provide adequate information. In more heterogeneous districts, comparisons might more appropriately be made between schools with comparable student populations and resources. Figure 1 could be easily modified to cluster together schools with roughly comparable populations and make this comparison visually compelling.

NYS Pupil Evaluation Program

Reading Grade 3

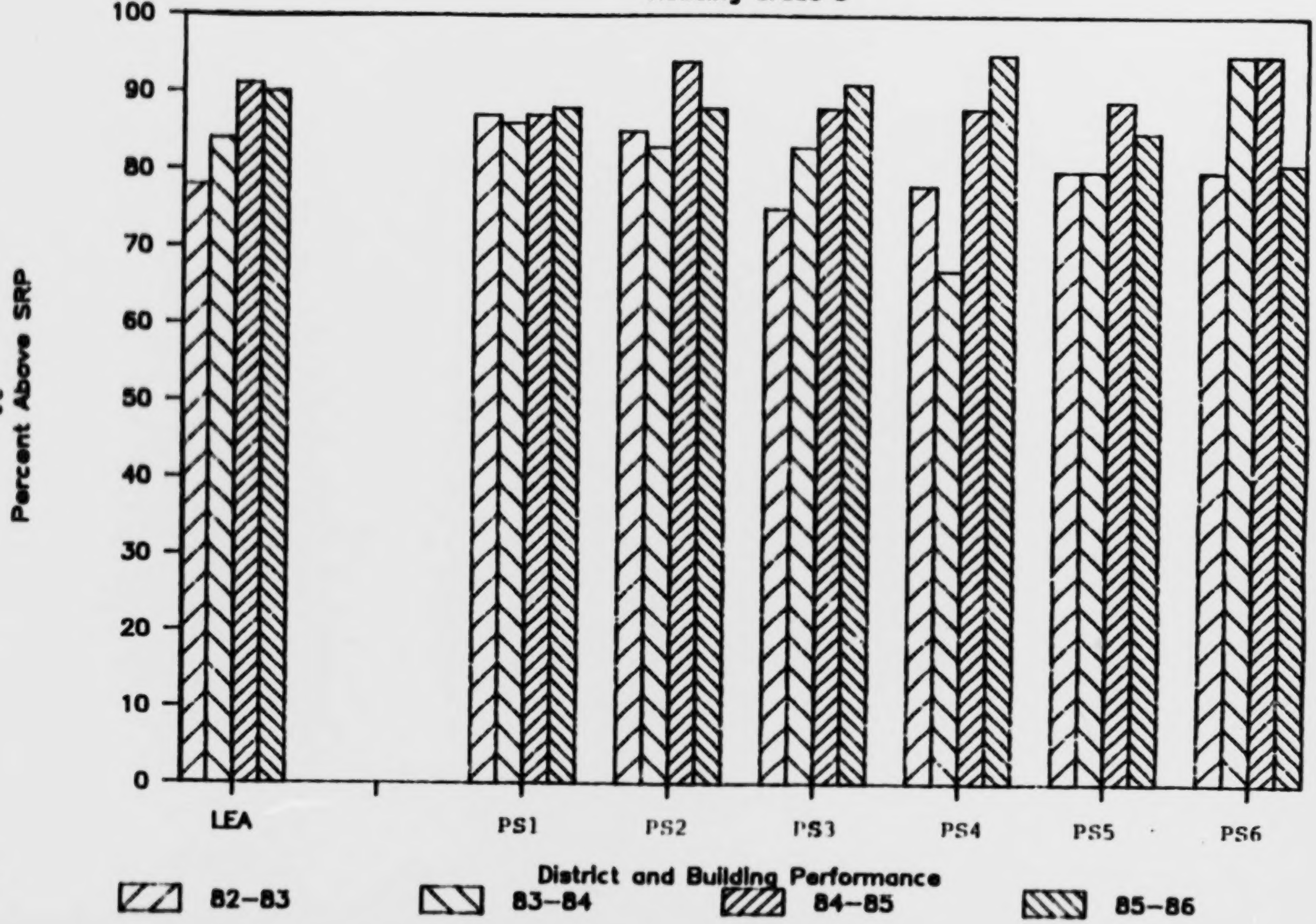


Figure 1

A next reasonable point of analysis for districts concerns performance across the basic skill areas. Are students doing better work in reading than they are in writing? Figure 2 on the next page shows reading and writing trend data for our hypothetical district and for each of its middle schools. One can see that it looks as if the district is doing somewhat better in reading. That being the case, before assuming that this variation in achievement is a function of variation in the quality of instruction, be sure to rule out the possibility that the difference is not a result of differences in who is taking the tests. Perhaps the more able writers are taking the English Regents rather than competency exam, and the effect springs from less able students being overly represented in the group taking the writing competency test.

Another important consideration for the district is grade-level performance. The CAR provides data for different groups of students over three years: third grade scores in 1983-84, third grade scores in 1984-85, and third grade scores in 1985-86. As a result, analysts often focus their comparisons on separate cohorts, a cohort being a grouping established by the year students entered a certain grade. The preoccupation with different cohorts tends to suggest that fluctuations in test scores are related to variation in the quality of teaching over time. In reality the higher test scores in the more successful years could simply be a function of the district's having enrolled more advantaged children those years.

Reasonable standards of comparison require data for contiguous grades, i.e. third graders in 1984-85, fourth graders in 1985-86, and fifth graders in 1986-87. A thoughtful inquiry about grade level performance might be expressed as follows: Are more students above criterion in the fifth grade in 1987 than were above criterion in the third grade in 1985? One must, however, be cautious here when interpreting the data. There is a tendency to assume that if fewer students are above grade level in the fifth grade than were above grade level in the third grade that the school district is less instructionally effective in the later grades. This may be the case, but one must first rule out the possibility that the population shifted over time. Remember that new students came into the school system and others left over those two years, meaning that we are not precisely comparing the 1984-85 third grade class to itself in 1986-87, when it is a fifth grade class. Differences in achievement could simply be a function of the changes in the population tested. Perhaps a disproportionately higher percentage of students with good test scores left the system over those two years, leaving a corresponding higher percentage of disadvantaged students to take the test in the fifth grade. Only when those possibilities are ruled out (e.g. through longitudinal analysis) can one talk meaningfully about educational trends.

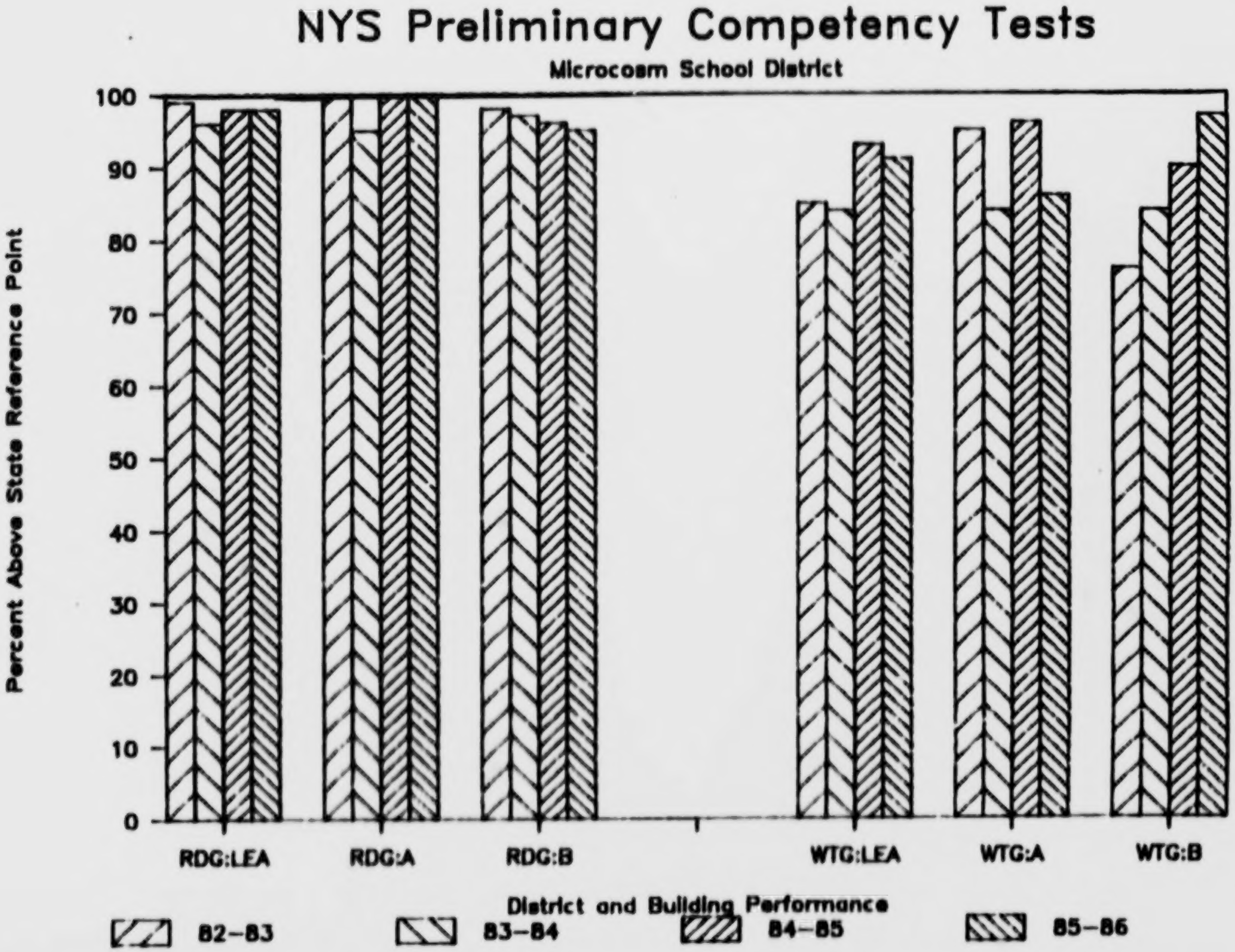

NYS Preliminary Competency Tests
Microcosm School District
Percent Above State Reference Point
100
90
80
70
60
50
40
30
20
10
0
RDG:LEA
RDG:A
RDG:B
WTG:LEA
WTG:A
WTG:B
District and Building Performance
82-83
83-84
84-85
85-86

Figure 2

School Level. Each school within the district would benefit from a similar scrutiny of the data, comparing their students' attainments across grade levels, across time, and across skills. The caveat regarding comparable populations still pertains when interpreting this information. To repeat an example given above, if the sixth grade in 1987 is not doing as well as it did when in was the third grade in 1984, the analyst should attempt to rule out the possibility that the effect is largely a function of mobility within the population.

To summarize, the logic of district level analysis is framed by the manner in which the data are presented by the state education department. In New York State, given the availability of school and grade level data across three time periods in several basic skill areas, it is appropriate, as has been suggested above, that comparisons be made on all those dimensions. The key to such analyses is the attempt to factor out the effects of extraneous variables and, to the extent possible, to emphasize comparisons among like groups.

Step Three: Supplementing the Data

During a press conference in November 1985 the Chancellor of New York City Schools announced that he would develop an assessment process that would go far beyond the requirements placed on the schools under the Comprehensive Assessment Report. The state's process, he said, had resulted in the erroneous impression that most city schools were miseducating youngsters. This section describes information New York City added to enhance understanding of the state's data and increase the value of those data for school improvement purposes. Alternatives that might be considered by suburban and rural areas also are included.

Academic Outcomes. New York City school profiles include achievement results from locally administered standardized tests as well as from the state achievement tests. New York City also includes the promotion rates from grade to grade for its schools. Other school districts may wish to supplement the CAR with criterion-referenced or diagnostic tests that can provide a detailed picture of content areas needing attention. Results of reimbursable program evaluation testing activities also may be included. Districts serving substantial numbers of academically oriented youngsters might catalogue students' awards, scholarships, SAT scores, and academic honors.

The purpose of this additional information is to help local policymakers more reliably judge if schools in their district are attaining their unique educational objectives. Since these are locally generated data, it is also possible for the district to report the information so as to detect underlying differences between the achievement of various student groups. Student achievement is known to vary by students' English proficiency, socioeconomic status, race, and gender to name a few conditions, and it is always interesting to see if schools are having any success reducing the predictive value of those variables. Although the New York State Comprehensive Assessment Report does not present data in a way that makes comparisons among such subgroups feasible, districts can attempt to organize their own local information gathering so as to facilitate such comparisons. To illustrate how this might be done, figure 3 on the following page compares male and female achievement on the College Board with nationwide results over the past three years.

Non-Academic Outcomes. While competency in reading, writing, and math are essential educational goals, there are many aspects of student growth and achievement that are as important, although in certain cases, more difficult to measure--creativity, verbal communication skills, perseverance, respect for one's own and other cultures, etc. The CAR does not report these student outcomes, but an individual district could as part of its assessment procedure. Accomplishments in areas such as music, art, community service, or crafts, if they are a critical component of a school or district's mission, could and should be highlighted. New York City, not unlike other troubled urban areas, is particularly concerned with reducing delinquency in its schools. New York City, therefore, records the number of unsafe incidents reported in its schools.

Institutional Characteristics. Although the quality of instruction is an important determinant of student success, research has demonstrated that there are a wide range of student characteristics associated with academic outcomes that are outside the schools' control. The New York State CAR contains information about some of these areas, for example, students' English proficiency, economic status, race, and mobility. The report also includes information about institutional variables that are associated with variation in student achievement. These include such matters as school size, class size, and pupil/support staff ratio, etc. Other variables that play a role in student progress and might profitably be measured include teachers' years of experience, the quality of instructional materials, extent of athletic facilities, etc.

In New York City, given the problem of urban overcrowding, the schools indicate their enrollment capacity and percent utilization. Suburban districts, where taxpayers directly determine the school budget, might have a special interest in specifying annual per capita expenditures for the three years in which there are outcome data. (It is difficult to imagine a school district - suburban or urban - that would not be interested in such data.)

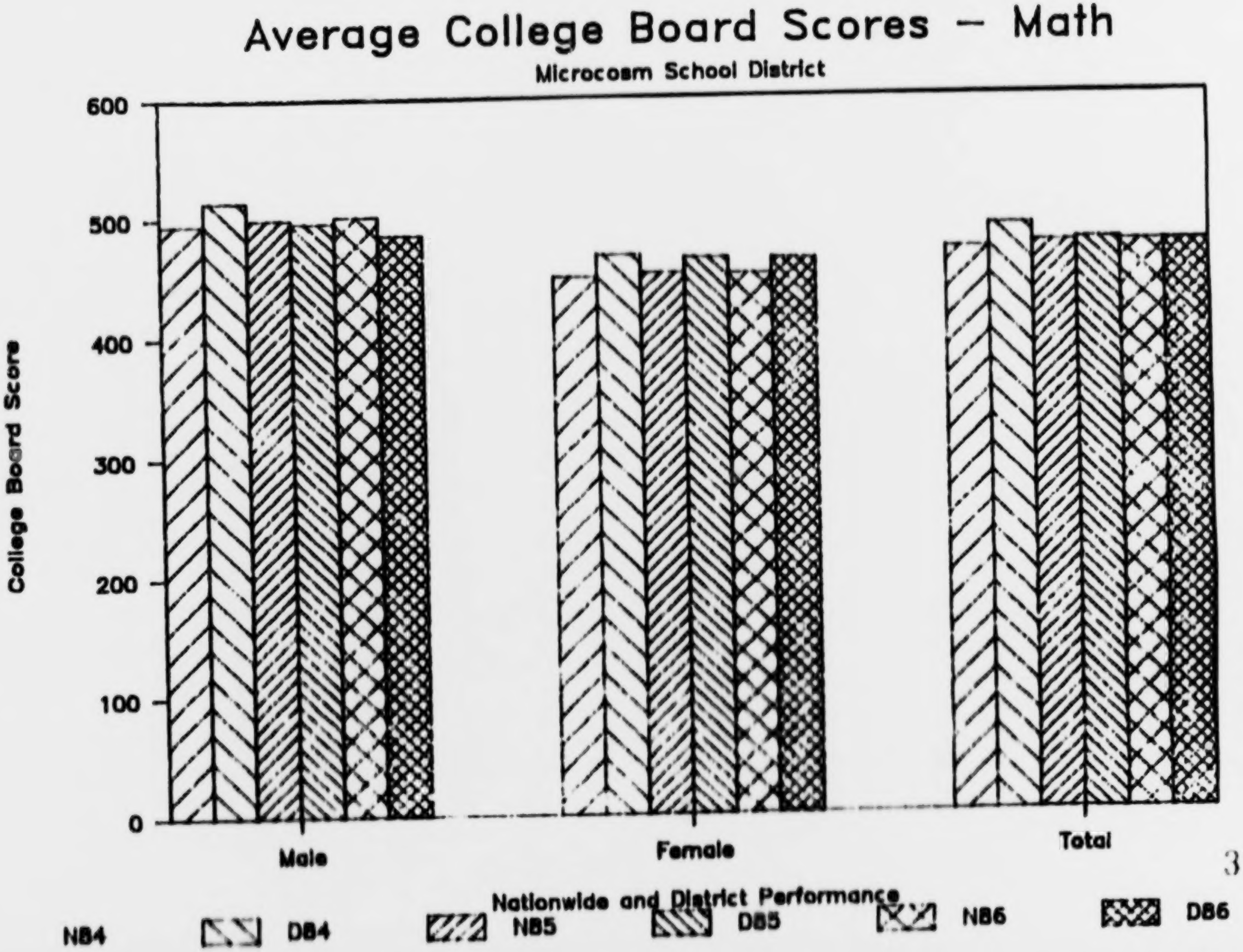

Average College Board Scores — Math
Microcosm School District
College Board Score
600
500
400
300
200
100
0
Male
Female
Total
Nationwide and District Performance
N84
D84
N85
D85
N86
D86

Figure 1

Step Four: Applying Appropriate Analytic Techniques

We have paid little attention so far to the application of inferential statistics to the data. It is our view that if districts carefully scrutinize the state's information, look for trends, and supplement the data with local information, that they will, for the most part, be able to isolate educationally meaningful patterns without advanced statistical techniques.

To illustrate an idiosyncracy of significance testing, if a school with 30 students in grade 6 had 60% of its students above the State Reference Point on the mathematics test in 1985 and 80% above in 1986, this apparently dramatic increase would not be statistically significant. The difference would have to be larger than 23 percentage points to be significant at the .05 level of confidence. When one is working with a small group (N=30), it takes a very large difference to produce a significant result.

What matters most is not whether a difference is statistically significant, i.e., that it is very unlikely to have occurred by chance. What matters most is whether there is enough general evidence in the data and in the "surround" to suggest an educationally meaningful trend.

While we are cautious about an over-reliance on statistics, there are often times when statistical techniques would be useful to confirm the significance of apparent trends in the data. We have, therefore, included a few examples of techniques that can readily be applied to the data presented by the New York State Education Department. Those techniques are described in Appendix 2. More sophisticated multivariate techniques, which allow one to estimate and control for the effects of several institutional and student variables simultaneously, will enhance understanding still further. Among the most useful multivariate techniques is multiple regression analysis, which makes it possible to estimate the degree to which certain factors are associated with student outcomes. An equation can then be developed that expresses that relationship. Working from the equation it is possible to identify "outliers," or schools that do better than others in their category, their category being schools with similar concentrations of the factors that are known to accompany achievement. For example, since socioeconomic status and mobility are related to academic achievement, a multiple regression analysis could allow us to identify schools that do better than expected after controlling for poverty and mobility, i.e., schools that do better than others within their socioeconomic and mobility groupings.

We have emphasized throughout this chapter that, when comparing group performance, comparisons be made with attention to the equivalence of the groups being studied. A multiple regression analysis simply gives us a statistical way of establishing this control. It is a powerful technique. Figure 4 on the next page illustrates one way in which the differences between schools' expected and actual scores might be presented. Schools scoring above the zero point have "residuals," or differences between expected and observed scores, that are higher than anticipated, and schools scoring below the zero point have residuals that are lower than would be expected.

Presenting the Assessment Report

Presentation of the comprehensive assessment report can provide the public with an opportunity to review student achievement and other data, and it can be a significant step toward involving parents and the public in local school effectiveness activities. Transforming the sterile tables and statistics into a meaningful public presentation is, therefore, of utmost importance.

Whatever the medium employed for reporting to the public, clarity is essential for audience understanding, report credibility, and user application. In this context, clarity refers to explicit, unencumbered and jargon-free narrative, illustrations, and descriptions. It is also characterized by conciseness, logical development, well-defined technical terms, tabular or graphic representations, and relevant examples (Joint Committee on Standards for Educational Evaluation, 1981).

An effective comprehensive assessment report should contain:

- a description of the assessment process;
- a summary of relevant demographics, so that an appropriate context for the analysis can be framed;
- a review of basic performance data, including state tests, graduation results, locally administered tests, and other outcome indices (it is recommended that data presentations should develop from the simplest descriptive statistics to any of the more complex comparisons and inferential statistics); and
- a summary of major findings.

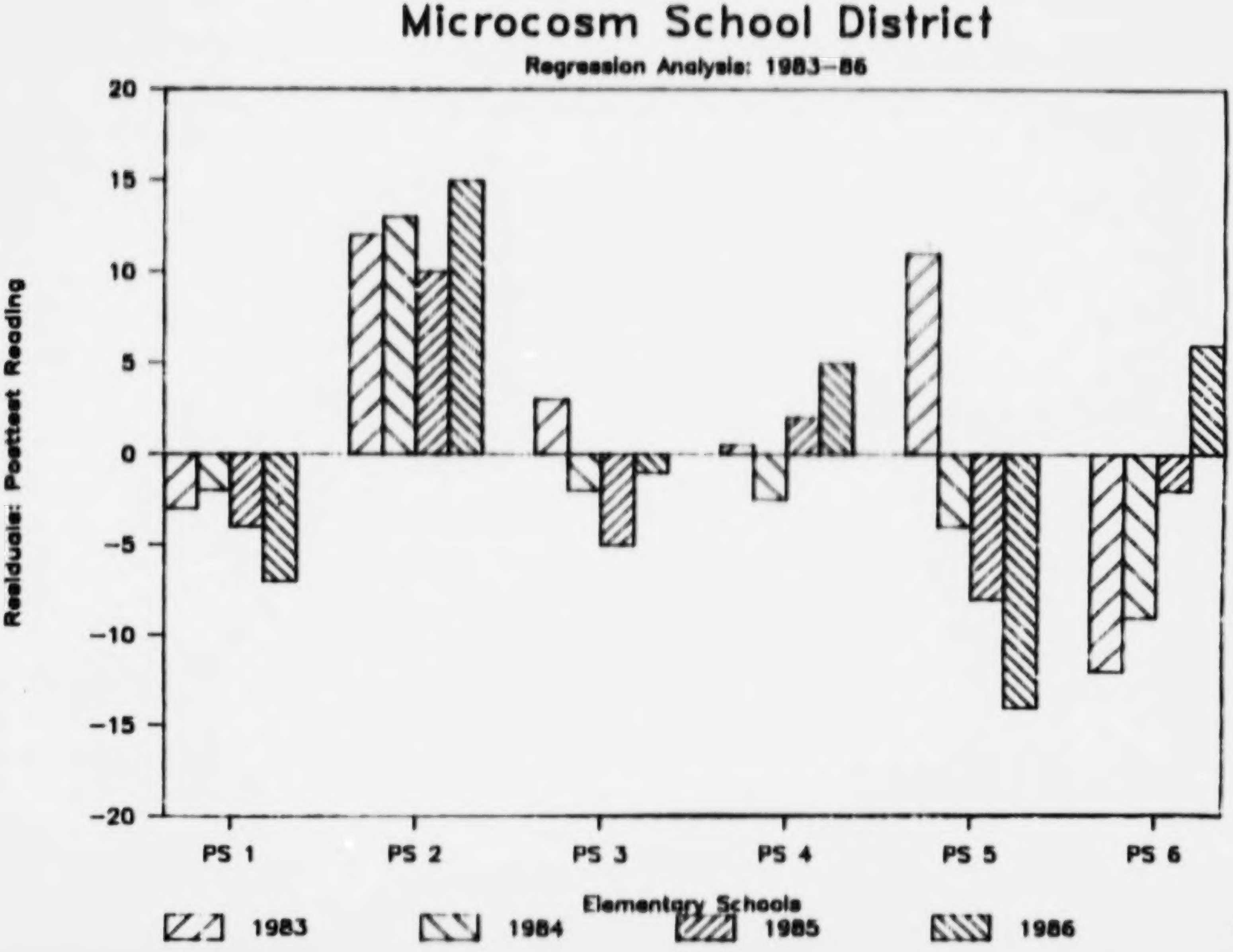

Microcosm School District
Regression Analysis: 1983–86
Residuals: Posttest Reading
20
15
10
5
0
-5
-10
-15
-20
PS 1
PS 2
PS 3
PS 4
PS 5
PS 6
Elementary Schools
1983
1984
1985
1986

Figure 4

The district's report presentation should be designed to stimulate public interest in collaborating on school level needs assessments and subsequent program development and evaluation activities. Included as Appendix 1 are excerpts from a model comprehensive assessment for a fictitious school district named Microcosm School District.

Microcosm School District is a mid-sized suburban school district with six elementary schools, two middle schools, and an academic high school. The district serves approximately 9,000 students in grades kindergarten through twelve. In addition to the standard testing program, Microcosm analyzes College Board exam data and annually administers several subtests of the Stanford Achievement Test (grades 1-9) and the Otis-Lennon School Ability Test (grades 2, 4, 6, 8 and 10).

APPENDICES

Appendix 1

MODEL COMPREHENSIVE ASSESSMENT REPORT: EXECUTIVE SUMMARY

***** Note to reader *****

A Model Comprehensive Assessment Report: Executive Summary. The following illustration has been compiled from handouts presented at December, 1986 public school board CAR presentations. In a typical presentation the Executive Summary is presented orally by the Superintendent (or his/her designee), while overhead transparencies are used to project the graphs. The following materials are normally distributed to the board members, district and school administrators, pedagogical personnel, parents, paraprofessionals, and others who attend the annual CAR presentation :

- the Executive Summary;
- the state's CAR Report;
- the state's guidelines for interpreting CAR data; and
- a locally prepared CAR Users' Guide to aid further with interpretation.

****

COMPREHENSIVE ASSESSMENT REPORT: EXECUTIVE SUMMARY
MICROCOSM SCHOOL DISTRICT

DECEMBER, 1986

A.B. PERSON, PH.D.
SUPERINTENDENT OF SCHOOLS

Microcosm School District
Comprehensive Assessment Report

Executive Summary

I. Introduction

A comprehensive assessment process was initiated last year by the New York State Board of Regents. The process began with the state's presentation of a Comprehensive Assessment Report (CAR), a document that summarized state test results, graduation data, enrollment and attendance data, ethnic distribution, class size, and other information about the total district and each school over three years. Districts were permitted to supplement the report with local information relevant to the assessment process.

The purpose of the Comprehensive Assessment Report was to provide a format for the Board of Education, parents, and members of the educational community to review and comment on the findings and to encourage these groups to work together to develop plans for school improvement.

The complete state report is available for review upon request. In some instances, the CAR developed by the state reported incorrect or missing data. To the extent possible, Microcosm School District has corrected such data and supplemented the report with additional information. This executive summary is designed to highlight the major findings of the CAR and to provide a concise description of the district.

This summary report is presented in the spirit in which the state intended, i.e., as a stimulus for discussion and planning and for increased involvement in and support for the educational programs of the district. It is important to remember that the report presents, for the most part, quantifiable information such as enrollment data, attendance data, test results and graduation data. Such results are, simply stated, the outcomes of the interaction of many factors. They should not be viewed as the "bottom line" that reflects the ability or performance level of Microcosm's students or teachers. The CAR utilizes quantifiable data to provide both a demographic and performance profile of Microcosm School District.

II. Demographic Data

An examination of the demographic data in the CAR shows Microcosm School District to be a mid-sized suburban school district. There are presently six elementary schools, two middle schools, and one academic high school in the district. During the past school year, approximately 8,575 students were enrolled in district schools. This represents a slight decline from the 9,047 students enrolled in the 1984-85 school year. Figure 1 shows elementary school enrollments for the district, as well as for each of the district's six elementary schools. It can be seen in the figure that, while there was an overall decline in enrollments, patterns varied from school to school.

Elementary School Enrollments

Microcosm School District

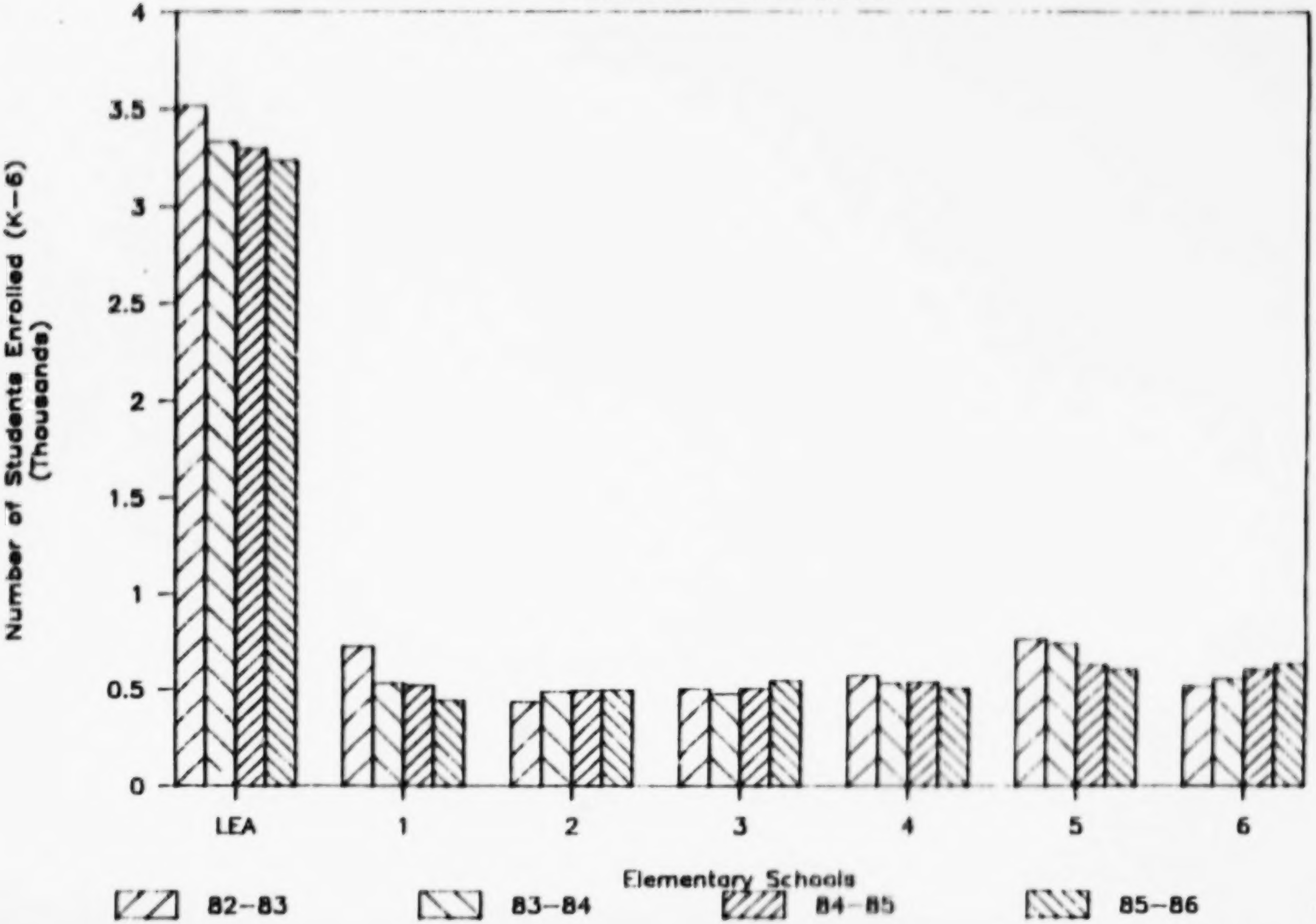


Figure 1

The annual attendance rate reported for the 1984-1985 school year for Microcosm School District is 93%. This rate has remained stable since the 1982-1983 school year. The dropout rate, which is the percentage of pupils who have left school for any reason except death or to enter another school, had increased slightly from 3.8% in 1983-84 to 4.3% in 1984-85. Attendance and dropout data for the 1985-1986 school year indicate that attendance has remained at about 93%, and that the percentages of students who dropped out of school or who have entered an alternative high school diploma or other diploma program have declined slightly.

With regard to ethnic distribution, Microcosm School District is primarily white (94.3%), with small percentages of students from Hispanic (3.0%), black (1.8%), and American Indian, Alaskan, Asian or Pacific Islander (.9%) backgrounds. Approximately 1.2% of the students enrolled are limited English proficient, i.e., they, by reason of foreign birth or ancestry, speak a language other than English and understand or speak little or no English. The CAR also reports socioeconomic indicators for the region. The 1980 Federal Census indicated that approximately 7% of the students in the district are members of families below the poverty level.

The Comprehensive Assessment Report presents class size information for classes in which enrollment is generally the largest: kindergarten, grades 1-6, grade 8 and grade 10 core subjects (English, math, science, and social studies). Microcosm's average class size, across all grades, is approximately 25 students. Each classroom teacher is, of course, supported by other subject area teachers, such as those who provide instruction in music, art, physical education, computer education, vocational education and library. A local study of the actual number of teachers in the district and the total pupil:teacher ratio supplements the average class size data presented in the CAR. Over the past three years, Microcosm School District has provided pupil:teacher ratios of 18 students to one teacher at the elementary school level and 15 students to one teacher at the secondary school level. These ratios reflect the above-mentioned support staff.

Taxpayers will be interested to know that, over the past several years, the tax rate has risen at a much slower rate than the district's operating expenses. The approved operating expense of the district, as represented by a cost per pupil, rose from $3,142/pupil in 1982-83 to $4046/pupil in 1985-86. This represents an increase of 28.8%. Despite these increases, the tax rate has increased by only 6.2% in the same period - from $45.64 per $100/AV in 1982-83 to $48.48 per 100/AV during 1985-86.

III. Performance Data

The Comprehensive Assessment Report focuses on quantifiable performance data such as state test results and high school graduation results in order to assess the district's past performance and current status. Last year the State Education Department examined such data for over 6000 schools in the state and identified the 600 schools with the poorest results in these areas. Such schools must develop and implement plans for improving their performance.

Not surprisingly, no schools in Microcosm School District were identified among those with the poorest results. The review of performance data for Microcosm indicates generally high levels of student achievement. The CAR developed by the state is supplemented in this summary with locally administered ability and achievement test data, including the Otis-Lennon School Ability Test, the Stanford Achievement Test and the College Board examinations. The review of these data enable the district better to prioritize needs and plan for program improvement. The analysis begun last year, when the CAR report provided achievement results for three years beginning in 1982-1983, has been extended into the 1985-86 school year. The four-year trends are summarized below.

A. State Tests

Test results are presented for four state testing programs: the Pupil Evaluation Program (PEP), Preliminary Competency Tests (PCTs), Regents Competency Tests (RCTs), and Regents examinations.

1. Pupil Evaluation Program (PEP)

The Pupil Evaluation Program assesses reading and mathematics in grades 3 and 6 and writing in grade 5. PEP tests are provided for use in the early identification of students who need special help in developing the basic skills of reading comprehension, mathematics, and writing.

Figure 2 shows four-year districtwide trend data on the PEP series for grades three, five and six. Figure 2a shows the same data for PS 01. Transparencies of Figures 2 and 2a are used to compare the school to the district's performance. By placing PS 01's transparency over the district's transparency it is easy to see discrepancies.

Figure 3 illustrates another approach to making school-to-district comparisons. In this figure third grade PEP reading performance is shown for the district and for each of its six elementary schools.

Across the four year period, Microcosm School District has demonstrated gains in the percent of students above the State Reference Point (SRP) in these skill areas. At the third grade level, 90% of the students were above the SRP in reading in 1985-86, whereas only 78% were above the SRP in 1982-83. Similar progress was made in third grade math, with the percent of students above the SRP increasing from 77% in 1982-83 to 98% in 1985-86.

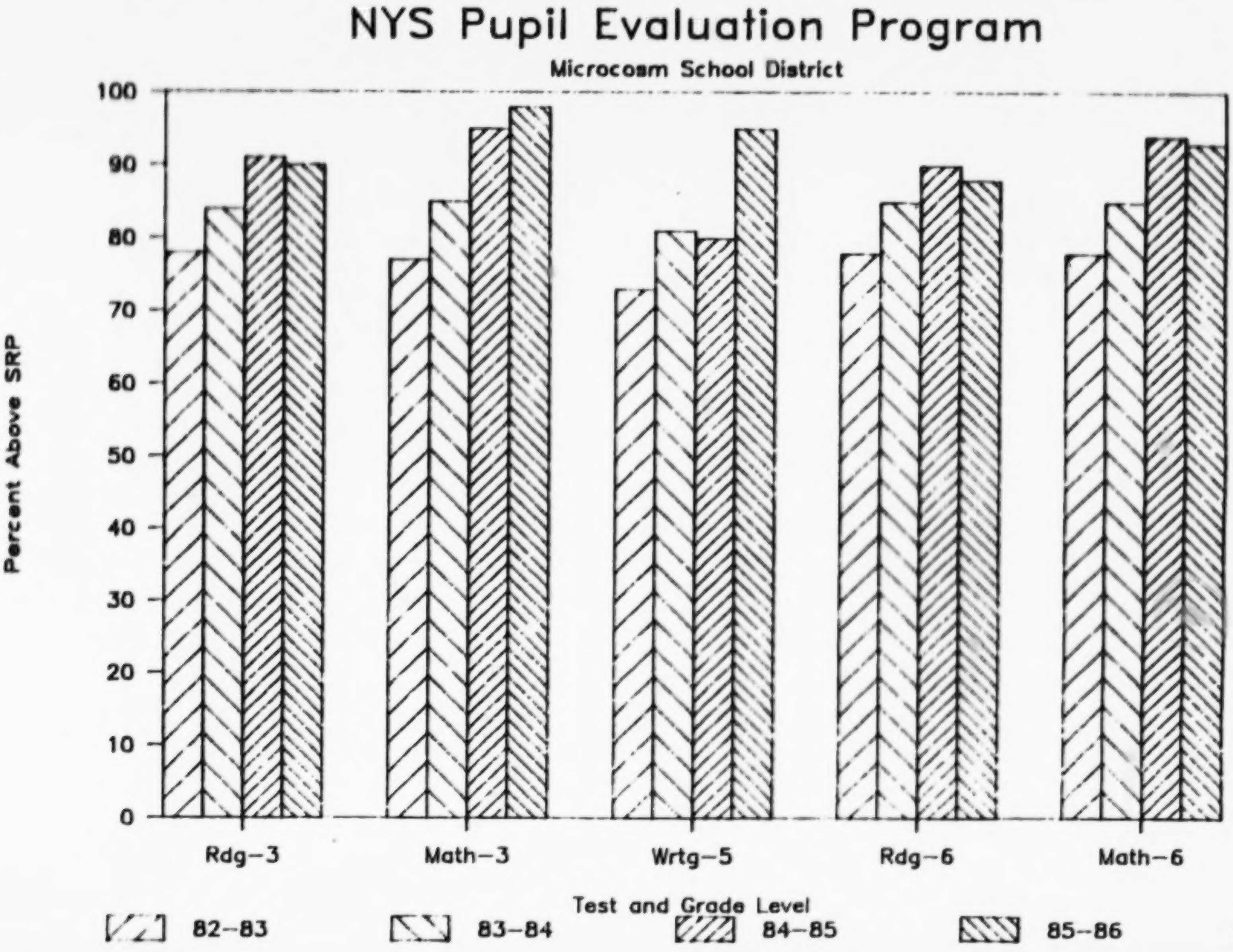

NYS Pupil Evaluation Program
Microcosm School District
Percent Above SRP
100
90
80
70
60
50
40
30
20
10
0
Rdg-3
Math-3
Wrtg-5
Rdg-6
Math-6
Test and Grade Level
▨ 82-83
▨ 83-84
▨ 84-85
▨ 85-86

Figure 2

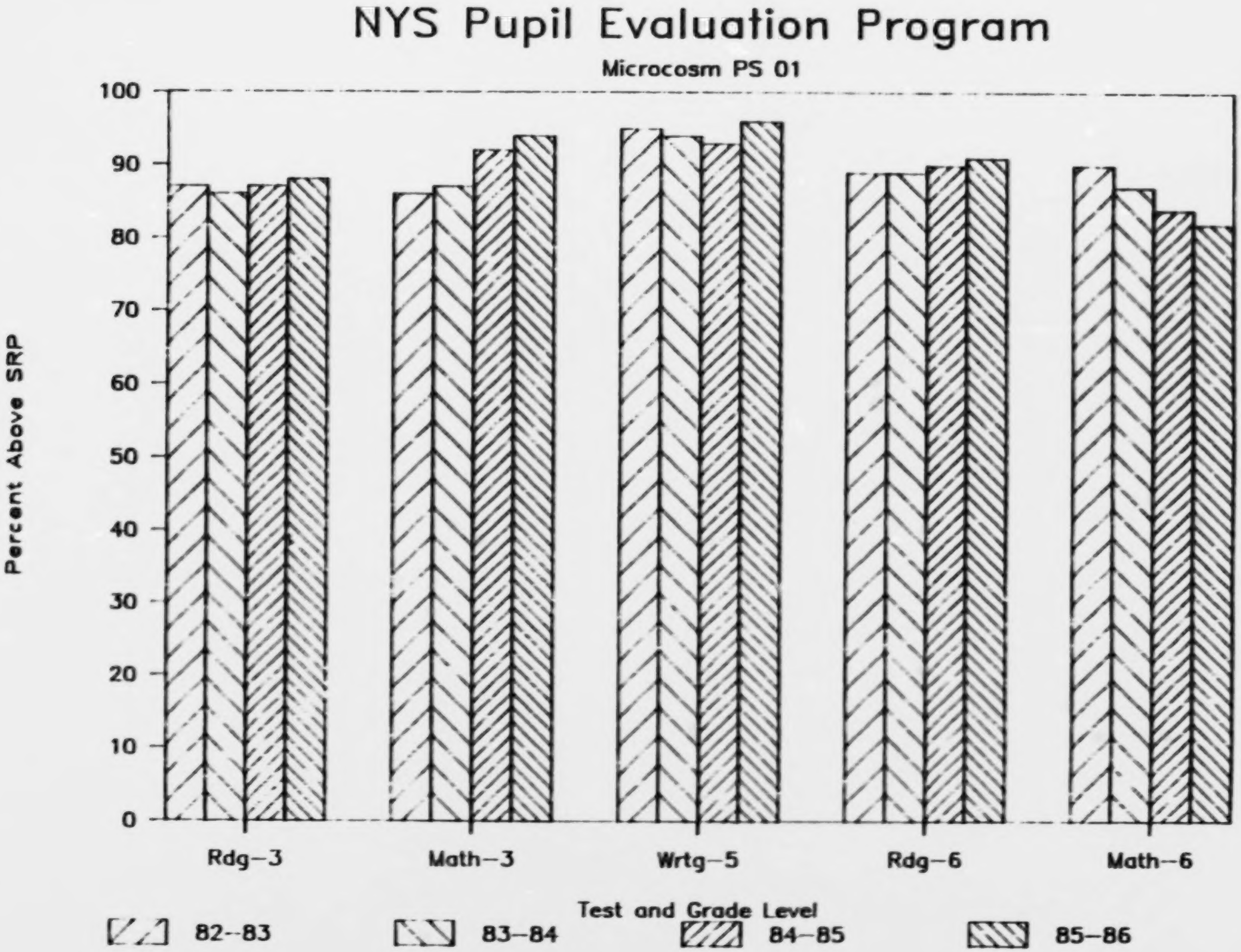
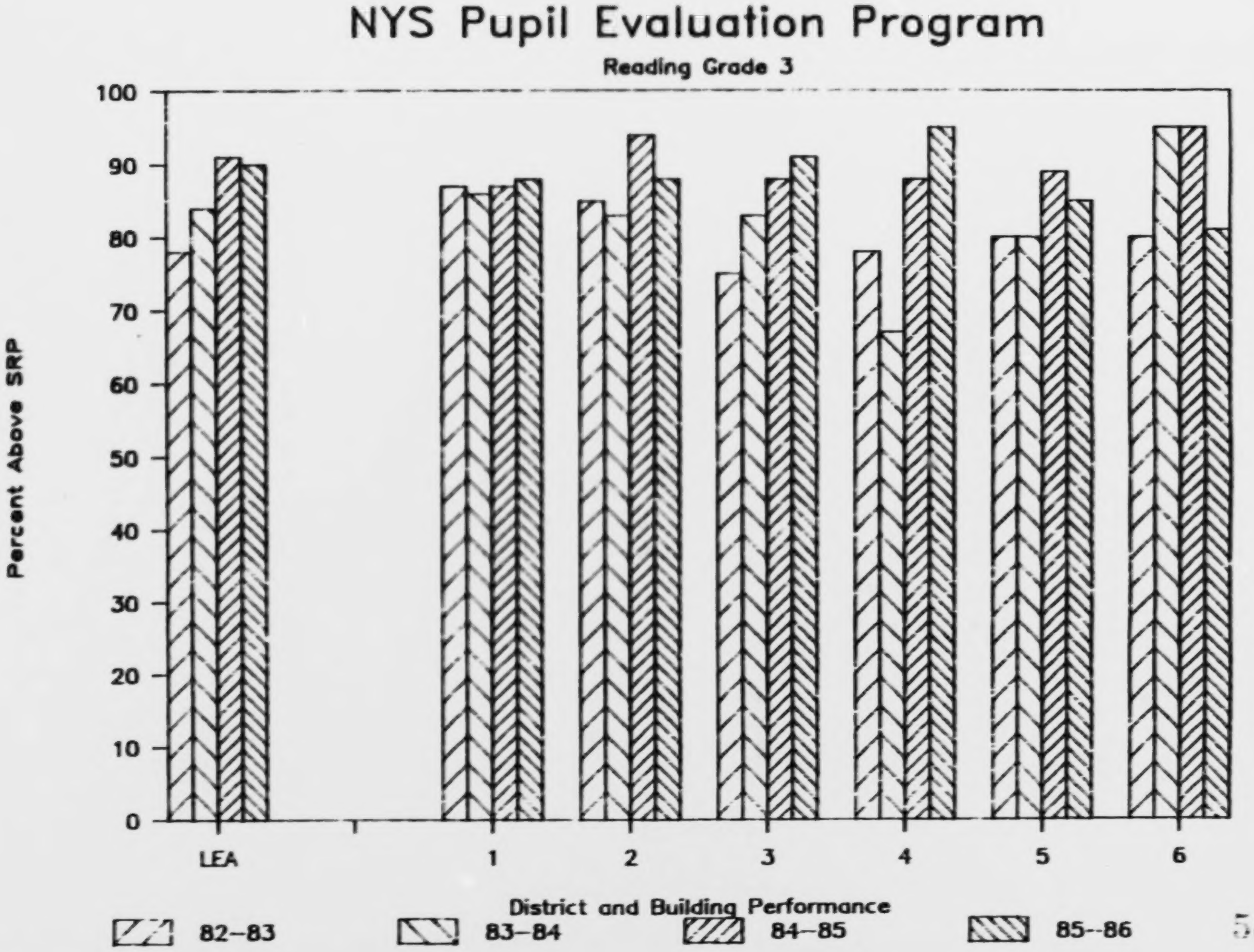

NYS Pupil Evaluation Program
Microcosm PS 01
100
90
80
70
60
50
40
30
20
10
0
Percent Above SRP
Rdg–3
Math–3
Wrtg–5
Rdg–6
Math–6
Test and Grade Level
82–83
83–84
84–85
85–86

Figure 2a



NYS Pupil Evaluation Program
Reading Grade 3
Percent Above SRP
100
90
80
70
60
50
40
30
20
10
0
LEA
1
2
3
4
5
6
District and Building Performance
82–83
83–84
84–85
85–86

Figure 3

At the sixth grade level, the percent of students above the SRP in reading increased from 78% in 1982-83 to 86% in 1985-86; for math, the percent increased from 78% in 1982-83 to 93% in 1985-86. Small declines in sixth grade reading and math from 1984-85 to 1985-86 are not statistically significant. In the area of writing, the percent of the fifth grade students above the SRP increased from 73% in 1982-83 to 95% in 1985-86.

2. Preliminary Competency Tests (PCT)

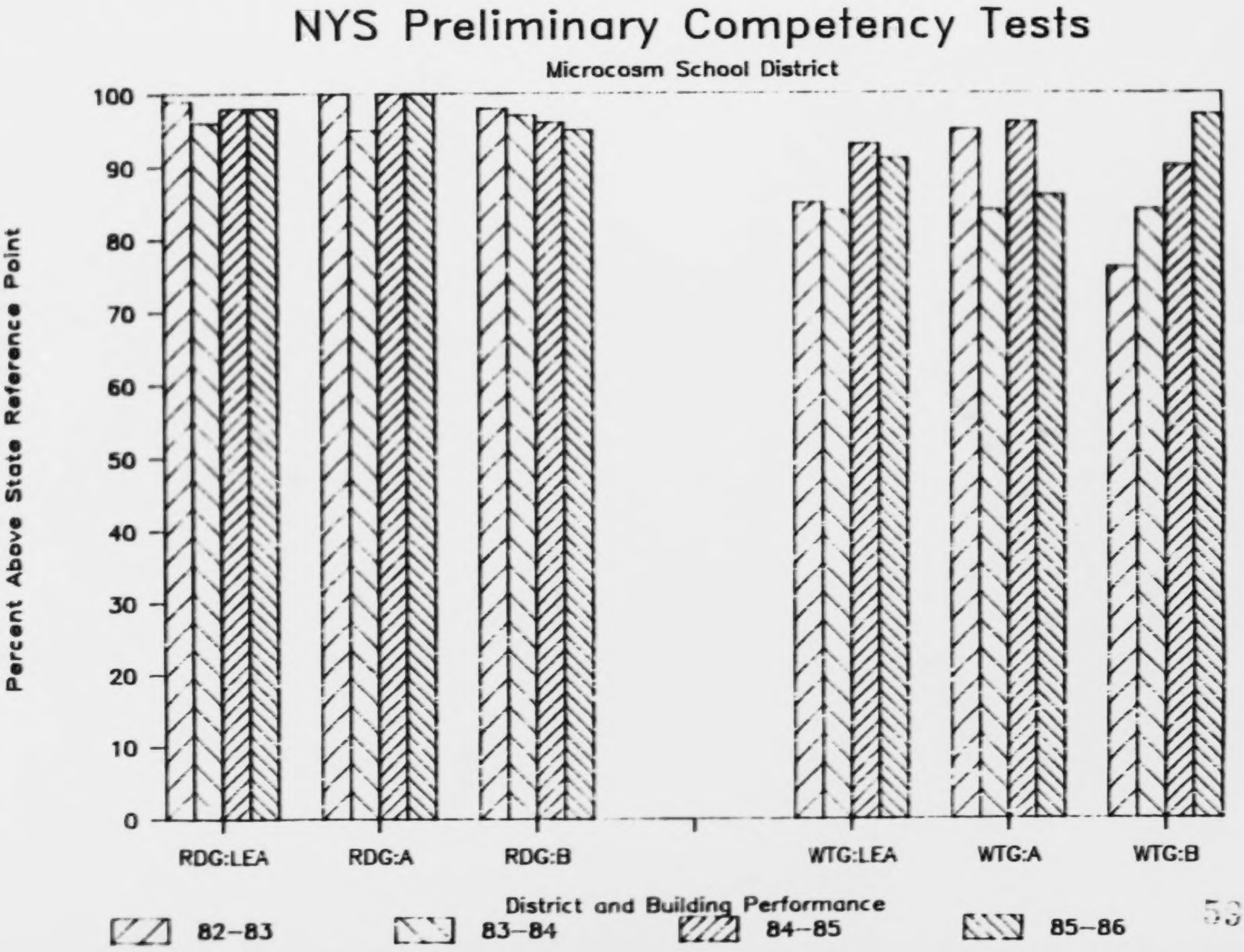
The Preliminary Competency Tests in reading and writing are given in either grade 8 or grade 9. Like the PEP tests, their primary purpose is the identification of students who need special help in developing the basic skills of reading comprehension and writing. Across the four-year period, Microcosm School District has demonstrated exceptional performance in the area of reading, with 96% of the students above the SRP. In writing, the percent of students above the SRP has increased from 85% in 1982-83 to 91% in 1985-86. It should be noted that, over the past four years, PCT reading performance at Middle School B has been showing a steady non-significant decline, while PCT writing achievement at the same school has had a dramatic and significant rise. Figure 4 shows PCT reading and writing trend data for the district and for each of its middle schools.

3. Regents Competency Tests (RCT)

In order to receive a local high school diploma, students must demonstrate competence in reading, writing and mathematics. The Regents Competency Tests (RCTs) provide one means for demonstrating the required level of proficiency. In most schools throughout the state, the RCTs are taken only by those students who are not planning to attend college and who do not take Regents examinations. Most college-bound students satisfy the competency requirements by passing Regents examinations.

Across the four-year period, the number of students in Microcosm schools taking Regents Competency Tests in January and June in all skills areas has decreased. Because the enrollment remained relatively stable during the four-year period, we may infer that the decrease in the number of students taking RCTs was due to an increase in the number of students satisfying the competency requirements by passing Regents examinations. The data also indicate that the percent of students passing the January RCT in math increased from 51% in 1982-83 to 59% in 1985-86. There was little change in the percent of students passing the January RCT in reading, with 95% passing in 1982-83 and 94% passing in 1985-86. The percent of students passing the January RCT in writing increased from 79% in 1982-83 to 82% in 1985-86.

For RCTs administered in June, the percent of students passing remained stable in math (69%). The percent of students passing the RCT in reading fluctuated from 90% in 1982-83 to 31% in 1983-84 to 81% in 1984-85 to 78% in 1985-86. These fluctuations may be due to the small number of students taking these exams (N < 42). In the area of writing, the percent of students passing the RCT increased dramatically over the four-year period from 50% in 1982-83 to 86% in 1985-86.



NYS Preliminary Competency Tests

Microcosm School District

Figure 4

4. Regents Examinations

Regents examinations are achievement tests based on the courses of study recommended by the State Education Department for use in grades 9-12. Regents examinations provide students, parents, counselors, teachers, administrators, college admissions officers and employers with objective and easily understood achievement information. Passing scores on the Regents examinations in English and mathematics satisfy all the competency testing requirements for a high school diploma. Figure 5 illustrates districtwide trend data for performance on several pairs of Regents examinations.

The CAR presents Regents examinations results for January and June administration periods. In general, few students took exams during the January examination period. For the June administration period, it is interesting to examine three pairs of tests that are typically taken as pairs by the same students: English and Social Studies, Biology and Math 10, Chemistry and Math 11. The data indicate that performance in biology, chemistry, and mathematics have generally declined in the past four years. The declines in biology and mathematics appear to be statistically significant. The percent of students passing English was always higher than the percent passing Social Studies; the percent passing Biology was always higher than the percent passing Math 10 (a new math Regents - Sequential Math II - replaced Math 10 and was administered in 1985-86); and the percent passing Chemistry was always higher than the percent passing Math 11. While there were fluctuations in the percent passing the Regents examinations in these areas across the four year period, it is important to note that the number and percent of students taking Regents examinations generally increased during the period. Across New York State, approximately 60% of the students in grades 9-12 take at least one Regents each year. In Microcosm, the percent taking at least one Regents each year increased from 59% in 1982-83 to 68% in 1984-85. This represents a 15% increase in the number of Microcosm School District students taking Regents. This datum was not provided by the State Education Department for 1985-86.

B. Graduation Data

The CAR provides information about the district's students who were candidates for graduation during each of the past three years. Figure 6 depicts Microcosm's high school graduation results relative to performance statewide. In the typical high school in New York State, approximately 90% of the diploma candidates receive a local high school diploma and approximately 40% receive a Regents diploma. For students who fail to graduate, the primary reason is failure to satisfy a local course requirement. In Microcosm, the percent of local high school diplomas has remained stable over the past four years - 96.8% in 1982-83 and 96.6% in 1985-86. Students in Microcosm demonstrate a higher graduation rate than those in New York State as a whole. It should be noted that there has been a steady decline in the percent of students who receive Regents diplomas in Microcosm, from 44.9% in 1982-83 to 35.8% in 1985-86 (4.2 percentage points). This decline will be discussed at upcoming district- and school-level planning meetings.

4. Regents Examinations

Regents examinations are achievement tests based on the courses of study recommended by the State Education Department for use in grades 9-12. Regents examinations provide students, parents, counselors, teachers, administrators, college admissions officers and employers with objective and easily understood achievement information. Passing scores on the Regents examinations in English and mathematics satisfy all the competency testing requirements for a high school diploma. Figure 5 illustrates districtwide trend data for performance on several pairs of Regents examinations.

The CAR presents Regents examinations results for January and June administration periods. In general, few students took exams during the January examination period. For the June administration period, it is interesting to examine three pairs of tests that are typically taken as pairs by the same students: English and Social Studies, Biology and Math 10, Chemistry and Math 11. The data indicate that performance in biology, chemistry, and mathematics have generally declined in the past four years. The declines in biology and mathematics appear to be statistically significant. The percent of students passing English was always higher than the percent passing Social Studies; the percent passing Biology was always higher than the percent passing Math 10 (a new math Regents - Sequential Math II - replaced Math 10 and was administered in 1985-86); and the percent passing Chemistry was always higher than the percent passing Math 11. While there were fluctuations in the percent passing the Regents examinations in these areas across the four year period, it is important to note that the number and percent of students taking Regents examinations generally increased during the period. Across New York State, approximately 60% of the students in grades 9-12 take at least one Regents each year. In Microcosm, the percent taking at least one Regents each year increased from 59% in 1982-83 to 68% in 1984-85. This represents a 15% increase in the number of Microcosm School District students taking Regents. This datum was not provided by the State Education Department for 1985-86.

B. Graduation Data

The CAR provides information about the district's students who were candidates for graduation during each of the past three years. Figure 6 depicts Microcosm's high school graduation results relative to performance statewide. In the typical high school in New York State, approximately 90% of the diploma candidates receive a local high school diploma and approximately 40% receive a Regents diploma. For students who fail to graduate, the primary reason is failure to satisfy a local course requirement. In Microcosm, the percent of local high school diplomas has remained stable over the past four years - 96.8% in 1982-83 and 96.6% in 1985-86. Students in Microcosm demonstrate a higher graduation rate than those in New York State as a whole. It should be noted that there has been a steady decline in the percent of students who receive Regents diplomas in Microcosm, from 44.9% in 1982-83 to 35.8% in 1985-86 (4.2 percentage points). This decline will be discussed at upcoming district- and school-level planning meetings.

4. Regents Examinations

Regents examinations are achievement tests based on the courses of study recommended by the State Education Department for use in grades 9-12. Regents examinations provide students, parents, counselors, teachers, administrators, college admissions officers and employers with objective and easily understood achievement information. Passing scores on the Regents examinations in English and mathematics satisfy all the competency testing requirements for a high school diploma. Figure 5 illustrates districtwide trend data for performance on several pairs of Regents examinations.

The CAR presents Regents examinations results for January and June administration periods. In general, few students took exams during the January examination period. For the June administration period, it is interesting to examine three pairs of tests that are typically taken as pairs by the same students: English and Social Studies, Biology and Math 10, Chemistry and Math 11. The data indicate that performance in biology, chemistry, and mathematics have generally declined in the past four years. The declines in biology and mathematics appear to be statistically significant. The percent of students passing English was always higher than the percent passing Social Studies; the percent passing Biology was always higher than the percent passing Math 10 (a new math Regents - Sequential Math II - replaced Math 10 and was administered in 1985-86); and the percent passing Chemistry was always higher than the percent passing Math 11. While there were fluctuations in the percent passing the Regents examinations in these areas across the four year period, it is important to note that the number and percent of students taking Regents examinations generally increased during the period. Across New York State, approximately 60% of the students in grades 9-12 take at least one Regents each year. In Microcosm, the percent taking at least one Regents each year increased from 59% in 1982-83 to 68% in 1984-85. This represents a 15% increase in the number of Microcosm School District students taking Regents. This datum was not provided by the State Education Department for 1985-86.

B. Graduation Data

The CAR provides information about the district's students who were candidates for graduation during each of the past three years. Figure 6 depicts Microcosm's high school graduation results relative to performance statewide. In the typical high school in New York State, approximately 90% of the diploma candidates receive a local high school diploma and approximately 40% receive a Regents diploma. For students who fail to graduate, the primary reason is failure to satisfy a local course requirement. In Microcosm, the percent of local high school diplomas has remained stable over the past four years - 96.8% in 1982-83 and 96.6% in 1985-86. Students in Microcosm demonstrate a higher graduation rate than those in New York State as a whole. It should be noted that there has been a steady decline in the percent of students who receive Regents diplomas in Microcosm, from 44.9% in 1982-83 to 35.8% in 1985-86 (4.2 percentage points). This decline will be discussed at upcoming district- and school-level planning meetings.

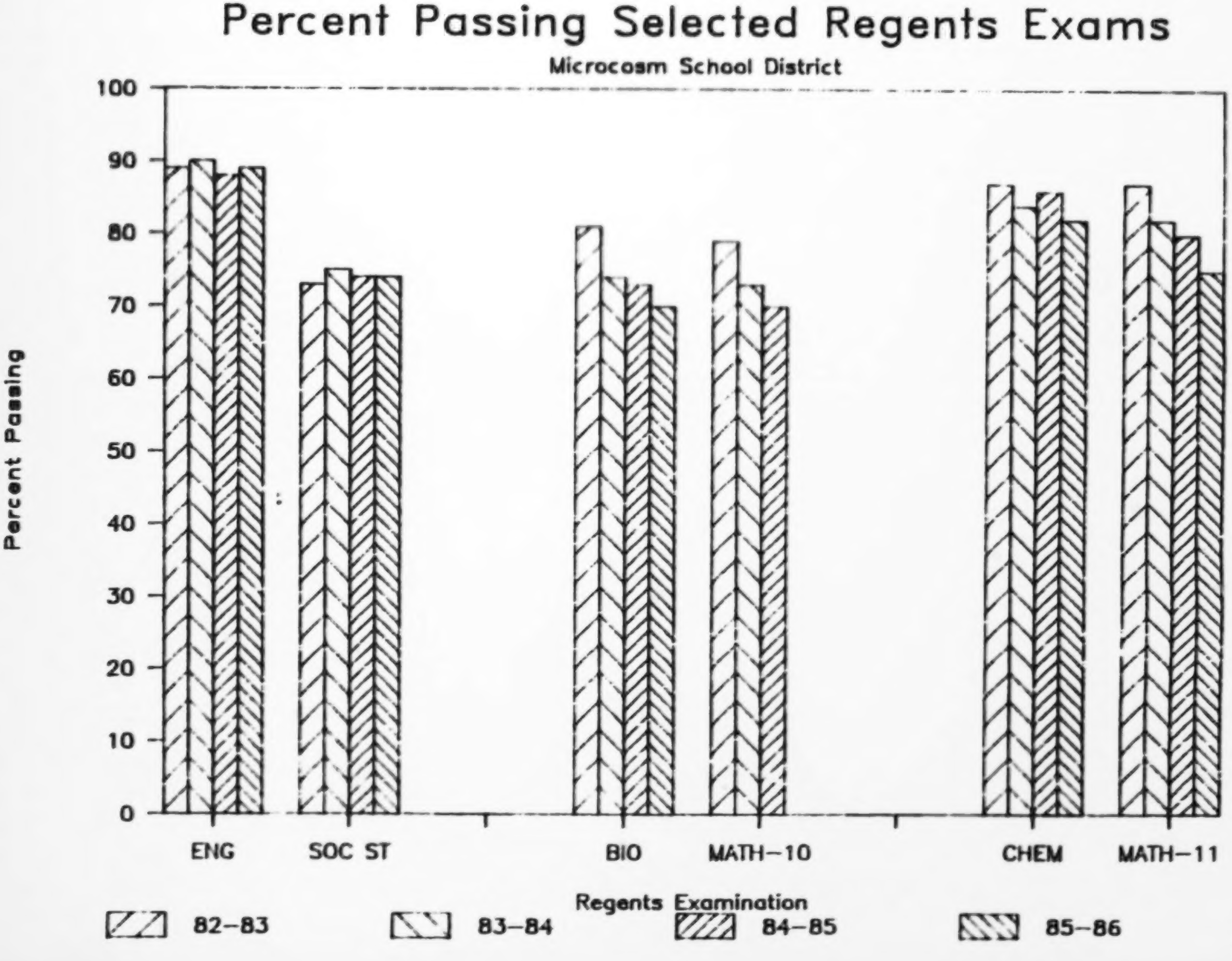
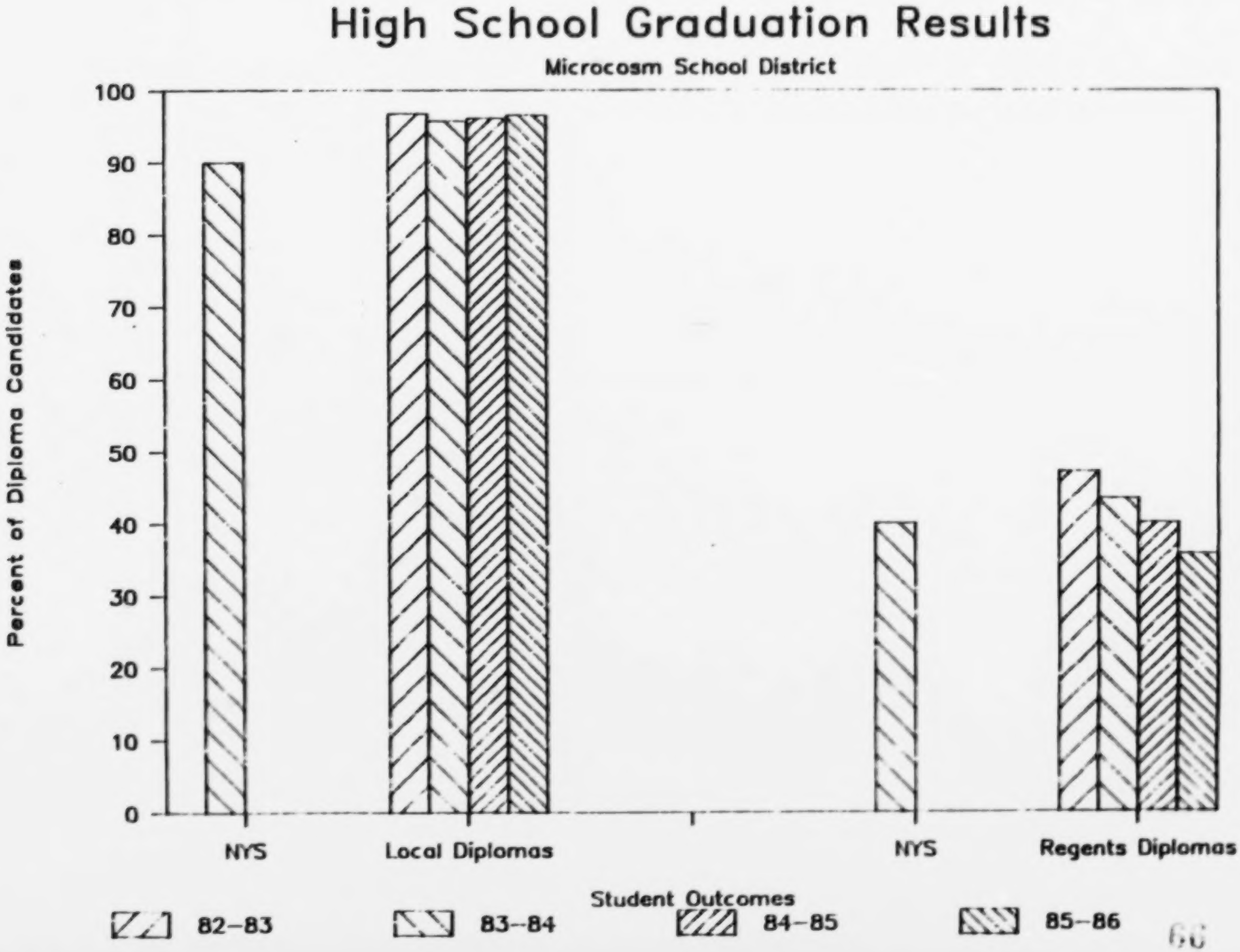


Figure 5

63

64


High School Graduation Results
Microcosm School District
Percent of Diploma Candidates
100
90
80
70
60
50
40
30
20
10
0
NYS
Local Diplomas
NYS
Regents Diplomas
Student Outcomes
82–83
83–84
84–85
85–86

C. Locally Administered Standardized Tests

In addition to administering the state tests, Microcosm School District also administers an extensive standardized achievement and aptitude testing program. The Stanford Achievement Test is administered to students in grades 1-9. The Otis-Lennon School Ability Test is administered to students in grades 2, 4, 6, 8 and 10. The College Board exams are administered to college-bound seniors. An examination of results from these testing programs provides additional information about the district.

1. College Board Exams

The College Board exams measure the scholastic aptitude of college-bound seniors in verbal and mathematical skill areas. In the verbal area, students nationwide showed improvements over the four-year period; average verbal scores rose slightly from 425 in 1982-83 to 431 in 1985-86. College-bound seniors in Microcosm demonstrated fluctuations in verbal scores across the same period, with their average verbal score varying from 418 in 1982-83 to 427 in 1983-84 to 417 in 1984-85 to 422 in 1985-86.

In the mathematical area, students nationwide also showed improvements across the four-year period; average math scores rose slightly from 468 in 1982-83 to 475 in 1985-86. College-bound seniors in Microcosm demonstrated fluctuation in math scores across the same period, with their average math score varying from 480 in 1982-83 to 475 in 1983-84 to 461 in 1984-85 to 460 in 1985-86. College Board mathematics data are illustrated in Figure 7.

Overall, the average performance of Microcosm School District students parallels that of students nationwide, with slight variations by sex across the years. (There is a slight decline among the boys and steady, but slightly lower performance among the girls.)

Despite a slight decline in the number of enrolled seniors, the number and percent of students taking the College Boards at Microcosm has increased over the four-year period.

2. Stanford Achievement Tests

The Stanford Achievement Tests measure student achievement in a wide range of skills, including reading, mathematics, and language. The Stanford Achievement Tests measure students' achievement in comparison to a representative national sample or norm group, and are thus referred to as norm-referenced tests. Norm-referenced achievement tests enable educators to compare the performance of an individual student or group of students to that of students in the norm group sample, which was selected to be representative of all major student populations in the country.

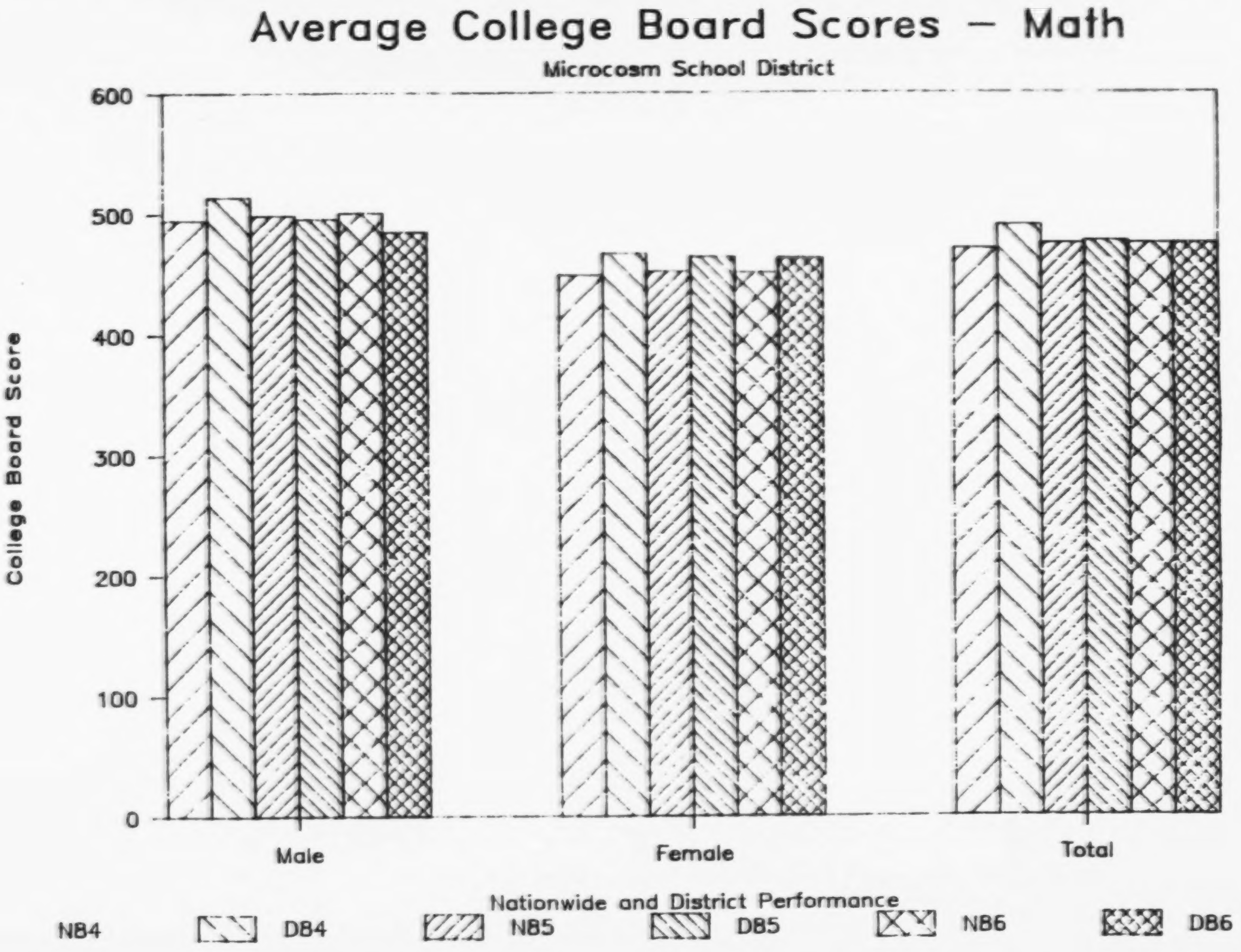

Average College Board Scores - Math
Microcosm School District
College Board Score
600
500
400
300
200
100
0
Male
Female
Total
Nationwide and District Performance
NB4
DB4
NB5
DB5
NB6
DB6

Figure 7

The Stanford Achievement Tests indicate the average percentile rank of Microcosm students among their peers nationwide. An examination of Microcosm's performance on the Stanford Achievement Tests in Reading and Math (grades 3 and 6) and in Language (grade 5) is provided in order to supplement the Pupil Evaluation Program data presented previously.

One way of examining performance on the Stanford is in terms of grade equivalent (G.E.) scores. Grade equivalents measure growth in reference to grade norms that are established by testing representative groups in each of a series of school grades. If students are tested in May (the 8th month of the school year), the expected grade equivalent score would be 3.8 for third grade, 4.8 for fourth grade, 5.8 for fifth grade, etc. The expected G.E. represents the mean of the national norm group. A G.E. score higher than the mean G.E. score for the grade indicates performance that is above that expected for a student at that grade level; a lower G.E. score indicates below average performance at that grade level.

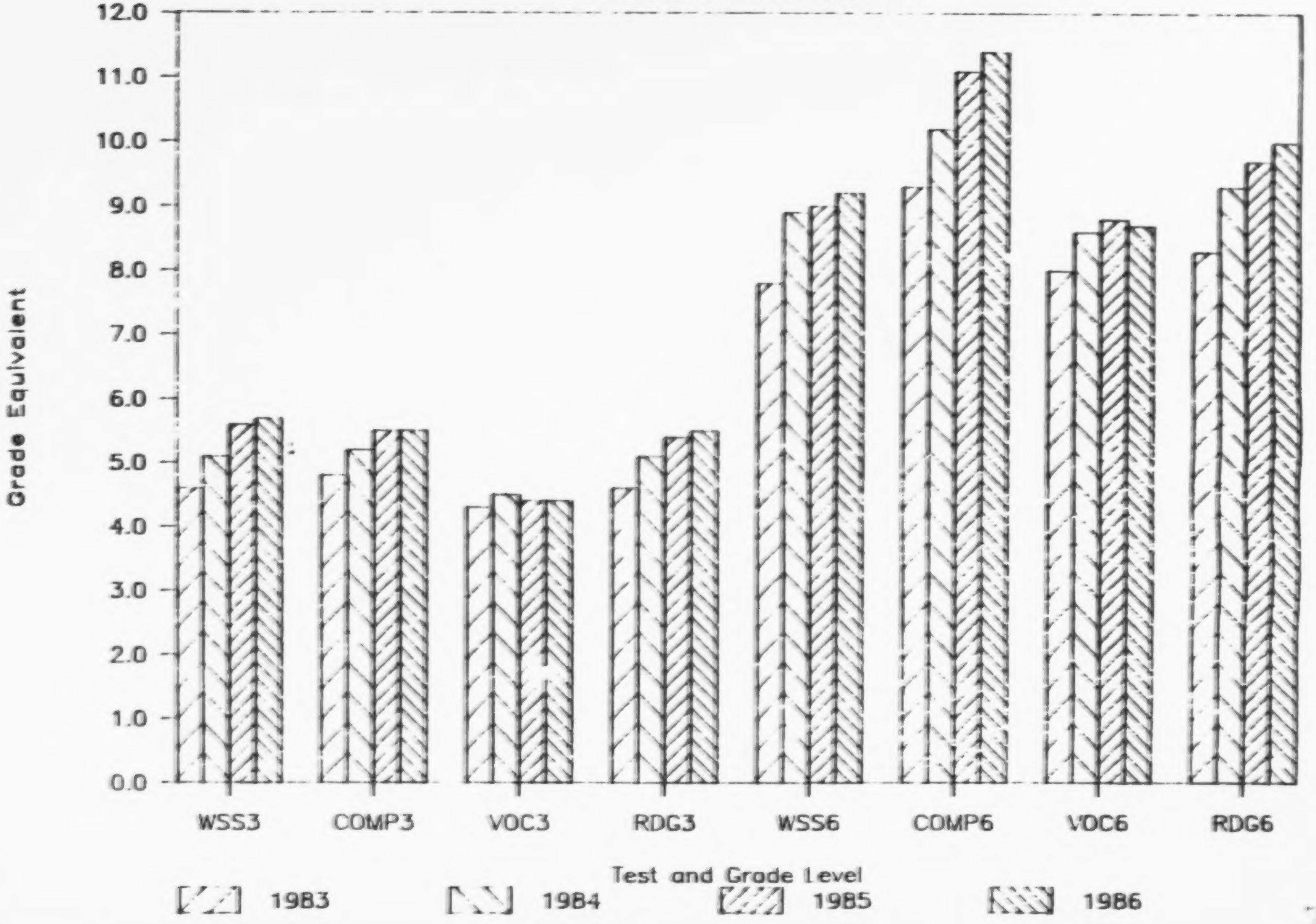
In the area of reading, Microcosm students showed generally rising G.E. scores in Total Reading, with the average G.E. rising from 4.6 in 1982-83 to 5.1 in 1985-86 for grade three, and from 8.3 in 1982-83 to 9.6 in 1985-86 for grade six. The expected G.E. was 3.8 for grade 3 and 6.8 for grade 6. At both grade levels, mean achievement was far above the expected level of performance for the grade. Figure 8 illustrates Stanford Achievement Test results for the various reading subtests (word study skills, comprehension, and vocabulary), as well as for total reading.

A similar pattern was shown in math, with the average G.E. scores increasing most years for both grade 3 and grade 6 in every subtest and in Total Math. In the area of language, increased mean G.E. scores were shown each year for grade 5 in Total Language.

Another way of looking at Stanford Achievement Test performance is in terms of percentile (%ile) ranks. A percentile rank indicates the percent of scores exceeded by the individual or group scoring at that rank. For example, a percentile rank of 85 means that students who attain the score corresponding to that rank exceed the scores of 85% of the students in the national norm group. The mean of the national norm group is the 50th percentile.

With the exception of third grade vocabulary, in the area of reading both third and sixth grades demonstrated increased average percentile ranks on every reading subtest. For example, mean comprehension scores increased for grade 3 from the 89th percentile in 1982-83 to the 95th percentile in 1985-86; for grade 6, from the 91st percentile in 1982-83 to the 96th percentile in 1985-86.

A similar pattern was shown for math, with increases since 1982-83 in the mean percentile rank for both grade 3 and grade 6 on every subtest. For example, mean applications scores increased for grade 3 from the 87th percentile in 1982-83 to the 96th percentile in 1985-86; for grade 6, from the 85th percentile in 1982-83 to the 97th percentile in 1985-86.



Mean G.E. on Stanford – Reading

Microcosm School District

Figure 8

For all language subtests, grade 5 students increased their percentile rank each year, with the average Total Language score rising from the 81st percentile in 1982-83 to the 94th percentile in 1985-86.

Stanford Achievement Test data also are used by Microcosm to evaluate the district's state-funded PSEN (Pupils with Special Educational Needs) diagnostic reading program. All participating students were pretested and posttested on appropriate levels of alternate forms of the Stanford Achievement Test in Reading. U.S. Office of Education Model A1 was used to analyze the data. A control group is not required in this design because the comparison is made between the student's expected progress without treatment (pretest normal curve equivalent) and his or her actual progress with treatment (posttest normal curve equivalent). Mean differences between pre- and posttest administrations were tested for statistical significance. Results indicate: significant gains in grades 4, 5, 7, 8 and 9; small but non-significant gains in grades 1 through 3; and a significant decline in grade six. Figure 9 illustrates these findings.

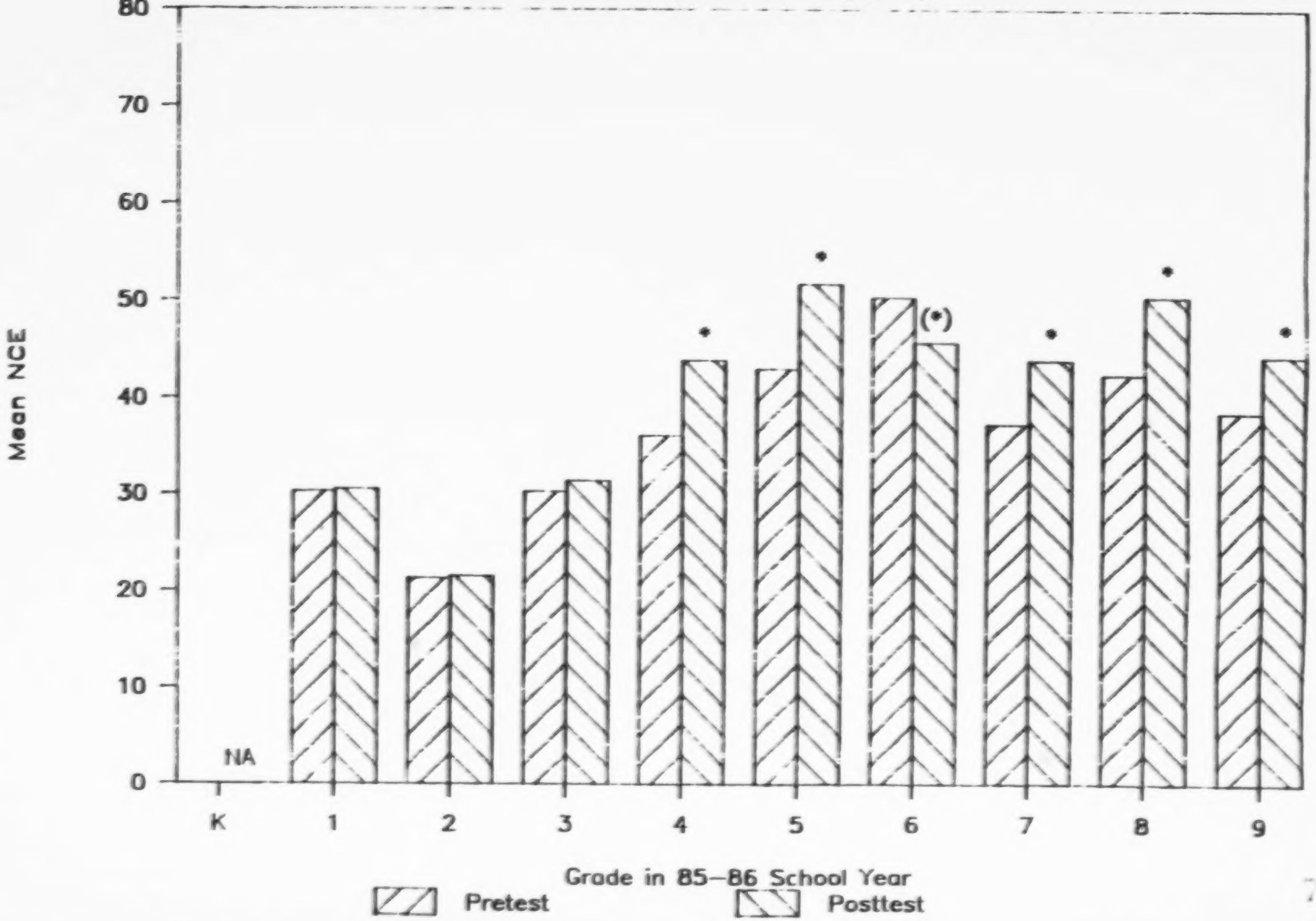
Finally, Stanford Achievement Test results are used by Microcosm in more sophisticated analyses of school effectiveness (i.e., multiple regression analyses). These analyses attempt to identify which schools are most effective in reading given the set of conditions that are known to impact upon achievement - e.g., poverty, mobility, attendance, pupil/teacher ratio, teachers' experience, etc. Effectively done they enable us to say which of Microcosm's schools are doing better (or worse) than expected, based on several of the schools' significant presenting conditions. Performance patterns for the past four years show PS 1 to be performing above expectation while PS 5's performance is below expectation. These data are shown in Figure 10. Implications of these findings are being reviewed with each school's planning team.

3. Otis-Lennon School Ability Test

The Otis-Lennon School Ability Test (OLSAT) is a power measure of abstract thinking and reasoning ability designed to predict success in cognitive, school-related activities. The OLSAT provides scores in terms of a School Ability Index (SAI: a measure of school ability similar to the IQ) and percentiles. An examination of Microcosm School District's OLSAT data indicates that, on the average, students in the district demonstrate school ability indices that are approximately one-half of one standard deviation above the norm. This means that Microcosm students are at the high end of the "average" ability range. The percentile ranks of the OLSAT serve as predictors of school performance. Using OLSAT percentiles, the predicted performance of Microcosm students would fall in the 62nd-79th percentile range. It is important to note that their actual performance is far higher than their predicted performance. This means that students in Microcosm School District are achieving at levels higher than would be expected from their OLSAT scores.

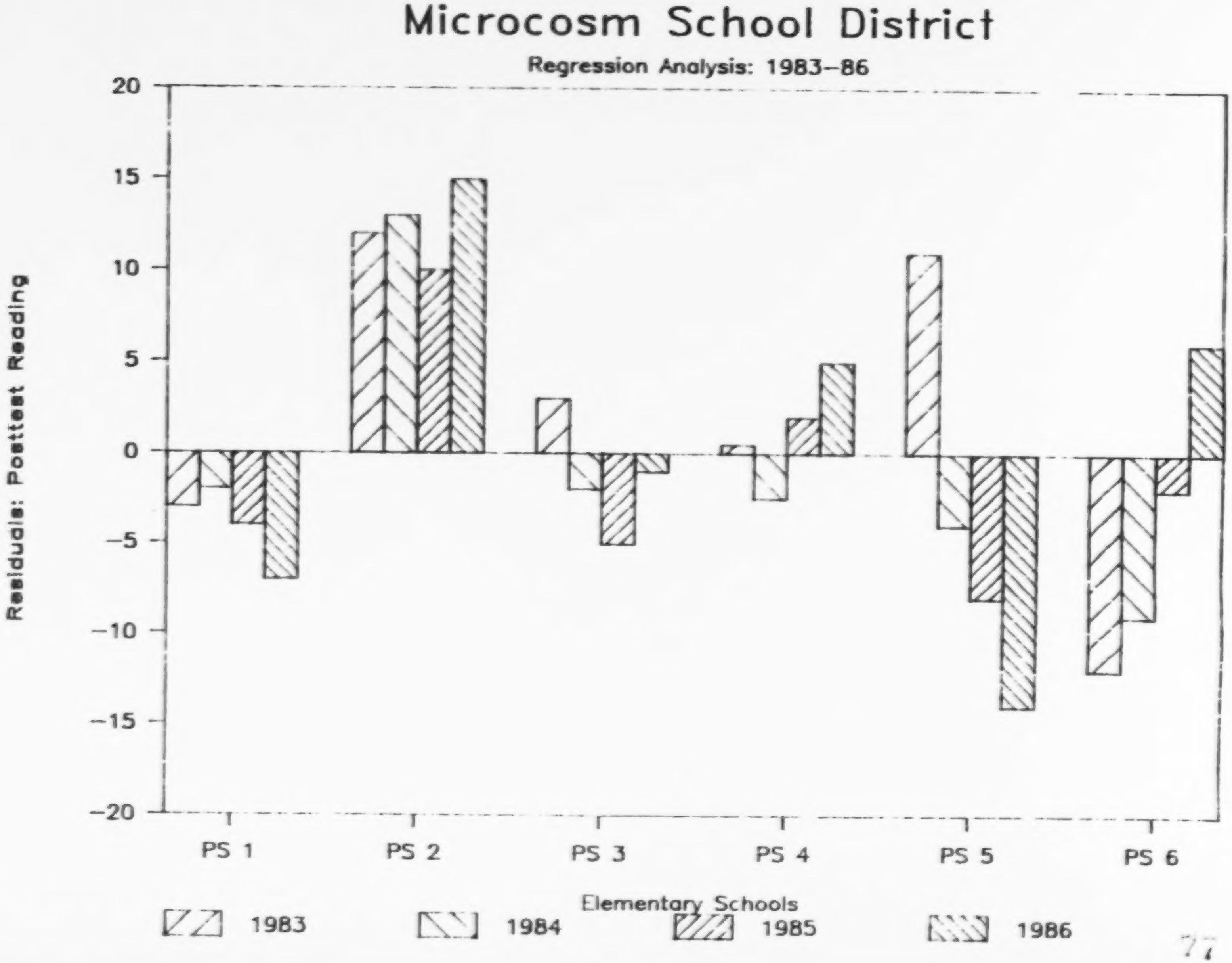
Microcosm School District

PSEN Diagnostic Reading (CAT)



*($p \angle .05$)

Figure 9



Microcosm School District

Regression Analysis: 1983–86

Figure 10

IV. Conclusion

The data presented in the CAR and supplemented by the district show Microcosm generally to be performing at a higher level now than four years ago:

- student performance has increased on state tests;
- more students are taking Regents exams; however there are significant declines in the percent of students passing biology and mathematics Regents and proportionately fewer graduates are receiving Regents diplomas;
- achievement of Microcosm students in comparison with their peers in a national norm group is higher than their expected grade levels, and average performance exceeds that of at least 90 percent of their peers nationwide;
- performance has increased steadily over the past four years on standardized, norm-referenced achievement tests; and
- reimbursable reading program (PSEN) participants in five of the nine grades demonstrate significant growth in reading, three grades show slight improvement, and one grade shows a significant decline.

These findings demonstrate that the district is on solid educational ground. Each school will utilize the data in the Comprehensive Assessment Report to: examine its past performance and current status; identify and prioritize need; and plan for school improvement. The district welcomes the continued involvement and support of the Board of Education, parents, and the community in planning programs that will effectively and efficiently address the needs, abilities and interests of all students in Microcosm School District.

Appendix 2

BASIC MEASUREMENT CONCEPTS AND ISSUES

The purpose of this appendix is to introduce some of the most basic measurement concepts and statistical techniques used in assessing school effectiveness and school change.

Understanding Normative Test Data

Test Construction

Can we use standardized tests to evaluate schools? Can we specifically use them to determine if schools are improving their capacity to educate youngsters over time? There is no completely satisfactory answer to that question. Understanding the dilemma requires a close examination of the basic attributes of tests, especially validity and reliability.

Validity. The most important characteristic of a test is its validity, which refers to the capacity of the test to measure what we want it to measure. There are three basic types of validity: (1) content validity, which measures the degree to which a test covers knowledge involved in a particular course of study, (2) criterion-related or empirical validity, which indicates how well a test predicts an individual's behavior in a specified situation (it is usually obtained through a correlation coefficient that expresses the tendency for values of the test to change systematically with the values of the criterion), and (3) construct validity, which refers to the extent to which a test may be said to measure generalized traits or theoretical constructs like intelligence and neuroticism that are presumed to inhere in individuals over time.

Reliability. Test reliability is necessary (but not sufficient) for good validity. By test reliability, we refer to the extent to which a test will rank order scores similarly for a group of people under differing conditions or situations. Note that it is reproducing the rank, not the same score, that is relevant. If everyone's scores were to change by the same amount under two conditions, the reliability of a test would be perfect.

Viewed from the perspective of reliability and validity, the central intellectual problem in test construction is to derive instruments in which the expected value of the indicators varies directly with variation in the construct or criterion and the rank ordering of individuals remains consistent. Standardized tests are not particularly effective instruments for measuring learning or change because they are constructed to yield stable rank orderings over time. Given this objective, items are selected for standardized tests that best differentiate pupils. Items that measure skills that are effectively taught to all students in school are not likely to be included in a test of that skill since they would have very little discriminatory power.

The question of using these instruments to measure growth, therefore, evolves into a discussion of whether it is possible to infer learning or cognitive growth from changes in students' relative positions on standardized tests despite the fact that the tests were not designed for that purpose. Understanding the issues involved requires an examination of the measurement properties of tests.

Levels of Measurement

The most critical aspect of standardized achievement tests that must be considered when using these instruments to assess learning regards the levels or scales of measurement and the interval properties of the scales. A scale is a scheme for the numerical representation of the values of a variable. There are four types of scales: nominal, ordinal, interval, and ratio.

Nominal Measurement. Many objects have characteristics that differ in kind only. Religion, occupation, and gender are examples of such variables--one is, for example, either male or female. In nominal level measurement, each value serves merely as a label or name. No assumptions are made about relations between the values assigned to the data.

Ordinal Measurement. When it is possible to rank categories according to some criterion, the ordinal level of measurement is applied. Ordinal measurement does imply a hierarchical ordering of values; however, the distance between any two values in an ordinal scale is unknown. It is therefore accurate to say that a score of 5 is higher than a score of 4, but how much higher is unknown. Additionally, the distance between 4 and 5 is not necessarily equal to the distance between 9 and 10. For example, the classification of enlisted personnel provides an ordering according to military rank, and each category has a clear position in the hierarchy. It should not be implied from this ordering that the difference between one rank and another is equivalent; distance is not measured.

Interval Measurement. With interval measurements the distance between values is meaningful. Unlike ordinal measurement, interval level data require the distance between one and two to be equal to the distance between 99 and 100, or any other set of scores. It is this property of interval scale data that permits one to add and subtract the values. The classic example of this type of scale is the thermometer, which measures temperature in degrees that are the same distance from each other at any point in the scale. The difference between 30 degrees and 31 degrees is the same as the difference between 8 degrees and 9 degrees. While an interval scale allows us to study differences, we cannot use it to discuss the proportionate magnitudes of those differences because an interval scale does not have an inherently determined zero point. We cannot say, therefore, that a temperature of 60 degrees is twice as hot as a temperature of 30 degrees.

Ratio Measurement. A zero point can be meaningfully designated with ratio measurements. It is, for example, possible to say that a car has traveled zero miles. Given an inherently determined zero point, it is possible to say that one car has traveled twice as many miles as another.

The meaning of these scales of measurement for the interpretation of test data is critical. Without interval measurement, growth rates in time or across groups of students cannot be measured. Some would say that because achievement tests are constructed to provide ordinal relationships, i.e., rank orderings of individuals, it is not strictly possible to subtract scores and arrive at a statement of differences that can be interpreted as learning. In response to this limitation, test construction experts have scaled test data to yield intervals. The rationale for imputing intervals is that the true values of the underlying construct are assumed to be normally distributed. Given interval level measurement, tests can be used to measure change across time since subtraction of scores is permitted. The discussion below provides some insights into these transformations of test data into equal interval data.

Score Types

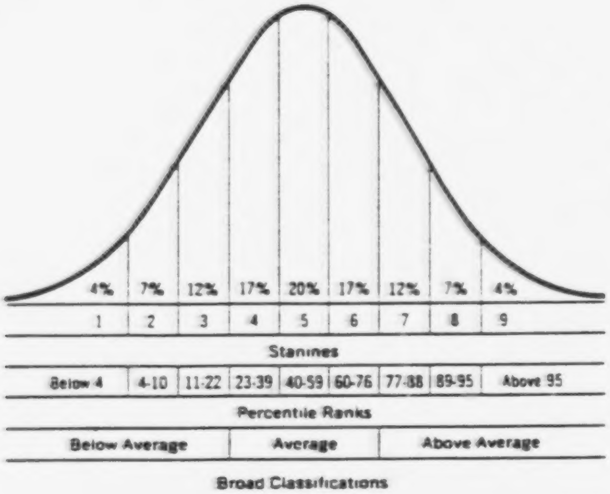
The problems using test scores to infer change or cognitive growth are not entirely resolved by the transformation of ordinal data into interval data. Each of the resulting score types has limitations. Those limitations are discussed below.

Raw Scores. The raw score is the number of questions answered correctly. For most users of test data, the raw score is of little value and must be translated into converted or derived scores.

Scaled Scores. Scaled scores express the results for a particular test and all forms and all battery levels - preprimer to advanced. While the scores allow one to see change over time, they have no inherent meaning or interpretive value. Percentile ranks, stanines, normal curve equivalents and, to a lesser extent, grade equivalent scores are preferable.

Percentile Ranks. A percentile rank indicates the relative standing of a pupil in comparison to the norm group. Percentile ranks range from a low of 1 to a high of 99 with 50 indicating typical performance. By way of example, if a pupil obtains a rank of 25 in reading, this means that 25% of the pupils in the norm group at the pupil's grade scored as well as or lower than the pupil; 75% scored higher. An important caution to keep in mind is that percentile scores tend to "bunch up" in the middle of the scale and spread out toward the ends, meaning that an increase in raw score points might translate into 20 percentile rank points in the middle of the scale and only 2 percentile rank points at either end of the scale.

Stanines. A stanine is a value on a nine point scale. There is a definite relationship between stanines and percentile ranks, with each stanine taking in a certain percentile band. Stanines, however, are equally spaced steps, and they do not bunch up or spread out at different points as percentile ranks do. The relationship between stanines and percentiles is illustrated below.



A Normal Distribution of Stanines, Percentile Ranks, and Performance Classifications

Grade Equivalents. A grade equivalent indicates the theoretical grade placement of pupils for whom a given score is typical. If, for example, a raw score of 35 corresponds to a grade equivalent of 4.9 it is interpreted that a score of 35 is typical for students in the ninth month of the fourth grade. Grade equivalents are losing their popularity, in part because lay persons frequently think that a pupil has mastered the curriculum of the grade corresponding to the grade equivalent obtained. When, for example, a student in the third grade obtains a grade equivalent of 5.4 in math, this is often misinterpreted to mean that the pupil can do the work taught in fifth grade. It is possible for a pupil to obtain a grade equivalent of 5.4 and have no ability to conduct mathematical operations typically learned in fifth grade. In reality the score indicates a performance as good as a fifth grade pupil's would be if that pupil had taken the third grade test.

Normal Curve Equivalents. This score is used almost exclusively for research purposes. NCEs are normalized standard scores with a mean of 50 and a standard deviation of 21.06. They range from 1.0 , which is the NCE corresponding to a percentile of 1.0, to 99.0, which is the NCE corresponding to a percentile of 99.0. The points between, however, are not equivalent. In comparison with percentiles, NCEs are more spread out toward the middle of the distribution and less spread out at the extremes. NCEs are the preferable metric in research, but many would argue that they do not have truly equal intervals. In essence, NCEs are simply a refinement on the stanine score; they share the same strengths and weaknesses as other normalized standard scores.

The description of score types illustrates that there are difficulties inherent in each of the measures. While NCEs provide a preferable metric to percentiles and grade equivalent units, they are not a panacea. They, too, are dependent upon ordinal data. For practical purposes, however, NCEs are treated as interval data.

The theoretical problems of assessing growth or learning from test data are, as we have seen, considerable. The statistics below are suggested to deal with these difficult problems. They are offered as a compromise between what is feasible to do in school settings and what can be expected to yield valid results.

Given the lack of consensus in measurement theory, the position taken in this guide is that when using tests to infer growth or change it is best to compare test results across the several available metrics and statistical techniques.

Using Statistics

The model comprehensive assessment report presented in Appendix 1 includes three common statistics for comparing aggregate outcome data: a test for the difference between proportions, a t-test for independent samples, and a t-test for correlated samples. The tests will be described briefly below following an introduction to basic statistical concepts. Individuals seeking a more complete explanation of the tests, their assumptions, and proper applications, or a consideration of alternative tests, may wish to consult a standard statistics text such as Statistics, 3 ed. by William Lee Hays (New York, New York: CBS College Publishing, 1981).

Basic Statistical Concepts

Measures of Central Tendency. The mean, median, and mode are frequently used to describe the central tendency of a distribution. The mode is the most frequently occurring value. The median is the number above or below which one half of the observations fall. If there are 29 observations, the median is the 15th largest observation. When there is an even number of observations, the median is the mean of the two middle observations. The mean is the arithmetic average or the sum of the value of all observations divided by the number of observations.

Measures of Dispersion. Two distributions can have the same measure of central tendency yet be very different. Consider two groups of three individuals in which the average age of the participants is 25 years. In the first group, each of the members is 25 years old, but in the second group one member is 50, another is 24 and another is one year old. There is more dissimilarity in these groups than there is similarity. A measure of dissimilarity or dispersion is needed to enhance our perception of the groups.

A commonly used measure is the variance. A variance is computed by summing the squared differences from the mean for all observations and then dividing by one less than the number of observations. The square root of the variance is called the standard deviation. The standard deviation has the appealing property of being expressed in the same units of measurement as the data itself, rather than in units squared as is the variance. The standard deviation of a distribution is the average absolute variation of all test scores around the mean value. For example, a standard deviation of 1.40 indicates "on average" test scores varied 1.4 points around the mean.

Standard Error of Measurement. The standard error of measurement indicates how much we would expect a person's score to vary if he or she were examined repeatedly with the same test. Since in any ordinary situation we have only one score for a person, we determine the standard error of measurement based on our knowledge of the normal curve and its properties. The score is derived by multiplying the standard deviation of the test by the square root of one minus the reliability coefficient for the test. The standard error of measurement is an extremely important score for test users. It allows us to set up confidence intervals or a range of score values within which, to a certain measurable limit, we can be confident that the true score falls.

Statistical Tests for Group Data

Test scores for individuals are notoriously unreliable. It is because of that unreliability that the standard error of measurement is such a critical concept for test users. Tests are far more useful when making judgments about groups because group data is a more stable measure of performance than scores for individual examinees. The following is a description of some of the most common statistical tests that can be performed with group data.

Differences between proportions. The numbers of youngsters passing the statewide reference point on standardized tests is the primary indicator of achievement in the New York State Comprehensive Assessment Report. Users of these data will, therefore, commonly want to know answers to questions such as the following: (1) Is there a difference between the percentage of third graders who passed the test reference point in 1986 and the numbers of third graders who passed the test reference point in 1987? (2) Is there a difference between the rate at which a certain cohort of youngsters passed the reference point in the third grade and sixth grade?

Solution: We test the difference between two proportions, P1 and P2, using the formula stated below. The symbol P1 represents the percent passing in group one; P2 represents the percent passing in group two. The symbol Q1 represents the result when P1 is subtracted from 1.00; Q2 is 1.00 minus P2. N1 refers to the number of subjects in group one and N2 to the number in group two. The symbol S is used to represent the standard deviation of the test statistic.

$$\text{Test Statistic } (Z) = \frac{(P1 - P2) - 0}{S(P1-P2)}, \text{ where}$$

$$S(P1-P2) = \sqrt{\frac{(P1)(Q1)}{N1} + \frac{(P2)(Q2)}{N2}}$$

For purposes of illustration, let us assume our interest is in determining if there is a difference between the rate of passing for two different, but highly comparable, groups of third graders; one group entered the system in 1986, the other group entered in 1987. We set our variable names as follows: P1 is the percent of third graders who passed the test in 1986 and P2 is the percent who passed the test in 1987. Again to illustrate the point, let us assume .40 passed the reference point in 1986 and .23 passed the reference point in 1987. Our hypothesis is that the difference between the proportions equals zero (H_0 : P1 - P2 = 0) or alternatively that the difference between the proportions does not equal zero (H_1 : P1 - P2≠ 0). Given the formula and the data for our problem, the test statistic is calculated as follows:

$$\text{Test Statistic (Z)} = \frac{(.40 - .23) - 0}{\sqrt{\frac{(.40)\,(.60)}{57} + \frac{(.27)\,(.73)}{57}}}$$

$$= 1.93$$

For a .05 level of significance and a two-tailed test, the decision rule is to reject the null hypothesis (H_0) if Z is greater than 1.96 or Z is less than -1.96. Since 1.93 is not greater than 1.96, we conclude there is no significant difference.

Two-Sample T-Test for Independent Means. The initial section of this appendix sought to demonstrate the theoretical problems encountered in attempting to determine differences between test scores with correlated samples. While there is some disagreement over the use of t-tests to calculate the difference between uncorrelated samples, most statisticians would admit of such a use. A typical problem requiring such an application occurs when districts are attempting to determine if, for example, the difference between the mean achievement of third grade students in 1986 and 1987 is significant. The means in this case are uncorrelated because the groups are composed of two distinct sets of students with no overlapping membership. The solution to this problem is given by the formula below.

$$\text{Test Statistic (T)} = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}}}\quad , \quad \text{where}$$

$$S_1^2 = \sum_{i=1}^{n} \frac{(Xi_1 - \bar{X}_1)^2}{N_1 - 1}$$

$$S_2^2 = \sum_{i=1}^{n} \frac{(X_i2 - \bar{X}_2)^2}{N_2 - 1}$$

The symbol X_1 represents the sample mean of a group, S^2, the variance, and N, the sample size.

Two Sample T-Test for Correlated Means. Districts may also be interested in knowing if the mean achievement of the same cohort varied between 1986 and 1987. In this case the means are correlated because the same students are in both groups. The formula is given below; however, we caution that the conditions for its application (high reliability of both tests and low correlation between the rank ordering of students on the two exams) are unlikely to apply, and the resulting information should be used cautiously. The symbols used are as follows: D is the mean difference between each of the paired scores and S_D is the standard deviation of that distribution; the other symbols are used as previously described.

Test Statistic (T for Correlated Samples) = $\frac{D}{S_D/\sqrt{N}}$, where

$$S_1 = \sqrt{\sum_{i=1}^{n} \frac{(Xi - \bar{X})^2}{N - 1}}$$

Appendix 3

Directory of School Improvement Resources

The annotated list of resources below is meant to supplement the discussion provided in the text. The selection includes books, manuals, and organizations that attend (although to varying degrees) to the concerns voiced in this guide. The primary focus of these resources, however, in not with the transformation of data into information, but the use of information for school improvement purposes, and as such they advance our concerns to their next logical step. For a comprehensive annotated listing of school improvement resources, we recommend readers consult Implementing School Improvement Plans: A Directory of Research-Based Tools (1987) published by The Regional Laboratory for Educational Improvement of the Northeast and Islands: Andover, Mass. The first selection we suggest is taken directly from that source.

Title:	Improving School Improvement: An Independent Evaluation of the California School Improvement Program
Author(s):	Paul Berman and Tom Gjelten
Description & Purpose	This report evaluates California's School Improvement Program (SIP), examining whether it has been successful and how it might be improved. Provides policymakers and citizens with basic information and understanding about how local instructional improvement can and does take place. SIP contains many elements of past education improvement efforts, but it also departs from those efforts in these ways: 1) program requires schools to adopt elements of a "model of change" focused on the change process, not on installing a particular innovation or instructional approach. Diversity in local objectives, plans and programs is promoted; 2) involves parents in actual program planning, changing the way decisions are made in schools; 3) intends to coordinate curriculum and instructional programs in a school with the student as

focus of planning efforts; 4) focuses on program quality rather than innovation and considers many outcomes in addition to student performance; and 5) State review provides formative feedback to the schools.

Consists of 10 chapters that cover these areas: assessment of SIP; the effects of implementation, school conditions, and districts on school improvement; implementation process and patterns; planning; school site councils and parent involvement; staff development; and program reviews.

Cost: free of charge

Contact Information:
Alex Law, Director
Program Evaluation and Research Division
California State Department of Education
P. O. Box 944272
Sacramento, CA 94244-2720
(916) 322-5010

Title: Building Effective Schools: Assessing, Planning, Implementing

Author(s): Janet Chrispeels and David Meaney, Ed.D.

Description & Purpose: Manual is designed to assist school administrators, principals, and other educational leaders who wish to implement a school improvement program based on the effective schools research.

Defines the term "effective school," describes characteristics associated with effective schools, and outlines the process being used by the San Diego County Office of Education to assist schools in becoming effective. The process is based on research on effective schools, effective classroom teaching, organizational development, and educational change.

Goal of the program is to increase the overall level of student achievement for all students through a process of assessment and planning and the implementation of a site-generated school effectiveness plan. Purposes are: 1) to determine the existence of specified elements needs to achieve school effectiveness; 2) to assess the school's academic effectiveness; 3) to determine the school's climate; 4) to establish a school action plan for improvement; and 5) to implement a plan for improvement.

Consists of four chapters and tables, charts, and appendices that include the following materials that could be useful staff development handouts: teacher/staff questionnaires, parent/student surveys, assessment instruments, sample comparisons, interview questions, evaluation forms, numerous test results, data collection forms, observation forms, worksheets and techniques.

Cost: $20.00

Contact Information:
San Diego County Office of Education
Room 212
6401 Linda Vista Road
San Diego, CA 94211-7399
(619) 292-3500

Title: All Our Kids Can Learn to Read: A Guide to Parent and Citizen Action

Author(s): Development by Chicago SCHOOLWATCH, a product of Designs for Change, Chicago, Illinois

Description & Purpose: Intended to be used as a handbook by parents, educators, citizens, and others who wish to take an active part in school improvement. Although there is an emphasis on improving reading skills, All Our Kids Can Learn to Read is very helpful in planning and implementing change in schools and can be applied to various subject areas and school improvement projects and issues.

Consists of five chapters that take an innovative approach to developing ingredients needs to teach/learn. Appendices include reading selections and principal interviews. Suggested program evaluation checklists and examples are included.

Chapter 1 is an introduction to SCHOOLWATCH and the Campaign for Effective Schools in Chicago, Illinois; Chapter 2, Are your Kids Learning to Read? analyzes skills needed to master the skill of reading (includes testing, evaluation, and the Chicago Mastery Learning Reading Program); Chapter 3, How Does YOur School Measure Up? describes ten ingredients identified as necessary for an effective school and includes the "SCHOOLWATCH Report Card" (checklist) that specifies what to look for when evaluating progress in school improvement; Chapter 4, What About the Bureaucracy? stresses decentralizing the bureaucracy and changes that can be made to improve how it works for the school; Chapter 5, Action, suggests active approaches to effective school improvement efforts.

Also available in Spanish.

Cost: $3.50

Contact Information:
Designs for Change
220 South State Street
Chicago, IL 60604
(312) 922-0317

Title: Instructional Leadership Handbook

Author(s): James W. Keefe and John M. Jenkins, Editors

Description & Purpose: Intended to be used as a reference book by principals working to become more effective instructional leaders.

The handbook is organized according to the role of the principal as instructional leader, that of providing direction, resources, and support to teachers and students for the improvement of teaching and learning in the school. The handbook reviews that in four areas: formative; planing; implementation; and evaluation.

Serves as a comprehensive resource for the principal who is learning to be a more effective instructional leader, and as a tool for use in staff inservice activities.

Formative Elements discusses trends in: content fields, organization and staffing, and media and methods; Planning Elements discusses 10 elements of effective planning, from needs assessment to budgeting for instructional improvement; Implementation Elements discusses one organizational and four supervisory elements: organizing the program, supervising classroom management, supervising the diagnostic process, supervising prescription and placement, and supervising instruction; Evaluation Elements lists six evaluatory elements: assessing and reporting student progress, teacher performance appraisal, program evaluation, community feedback, program modification and revision, and communicating the program.

Cost: $8.00

Contact Information:
Publications Department
National Assn. Secondary School Principals
1904 Association Drive
Reston, VA 22091
(703) 860-0200

Title: A Process Guide for School Improvement

Author(s): Herbert J. Klausmeier

Description & Purpose: Guide is for use in locally conducted inservice programs and in university courses and programs. It is designed to aid individual schools as well as district offices in starting and maintaining a self-improvement capability that is characterized by: 1) maintaining satisfactory student outcomes and improving unsatisfactory ones; 2) maintaining or improving effective instructional practices; 3) maintaining or improving job satisfaction and morale; and 4) increasing competency and professional development of staff. Chapter 1 is for all levels of schools, Chapters 2 through 11 are more for the middle school and high school, and Chapter 12 is more the elementary school. Chapters 2 through 11 give improvement suggestions and illustrative exemplary practices of both middle schools and high schools. Chapter 12 does the same for elementary schools. Most information regarding middle schools or elementary schools is relevant to both.

Each chapter includes an improvement plan, covering such areas as: administration/structural organization; program planning; curriculum planning; school climate; testing/data gathering/evaluation; teacher-advisor program planning; home/school/community relations; and components contributing to effectiveness and activity planning to increase effectiveness. In addition, the Appendix offers suggestions for using the guide in locally conducted inservice/staff development programs and supplies supplementary materials for secondary and elementary schooling. Also, simulations are available from the Wisconsin Center for Education Research (one each -- elementary, middle and high school).

Cost: $12.75 paperback, $28.00 hardcover

Contact Information:
University Press of America
4720 Boston Way
Lanham, MD 20706
(301) 459-3366

Title: Reaching for Excellence: An Effective Schools Sourcebook

Author(s): Regina M. J. Kyle, Ed.

Description & Purpose: An integrated document that contains summaries of the knowledge base as well as a current directory of programs and sources of technical assistance relevant to research and practical information in the area of school effectiveness/improvement programs being implemented in the United States. Intended for use by administrators, curriculum specialists, staff developers, and teachers.

The overview of the book emphasizes key themes and issues across the chapters on research that appear in Part I, including an extended discussion of staff development, and addresses the process of translating research into policy with illustrations.

Six of the seven chapters summarize the research from these perspectives: effective classroom practices in elementary schools and secondary schools; effective school practices at each level; district- and state-level practices that support effective school management and instruction. Includes a chapter on criteria and methods for measuring effectiveness. Part II presents a directory of successful effective schools programs being currently implemented.

Materials: Directory of Programs Promoting Effective Practices at the Classroom and Building Levels

Cost: $9.50

Contact Information:
Superintendent of Documents
U. S. Government Printing Office
Washington, D.C. 20402
(202) 783-3238

Title: The Practice Profile: An All-Purpose Tool for Program Communication, Staff Development, Evaluation, and Improvement

Author(s): Susan Loucks-Horsley and David P. Crandall

Description & Purpose: Provides a standardized, systematic, and cost-effective way to summarize program components and requirements. Provides guidance as to implementing and evaluating the practice and facilitating comparison with other programs. Especially helpful to program/project directors, staff developers, decisionmakers, evaluators, potential adopters, and researchers.

Description: Describes the three parts of a Practice Profile: a component checklist contains components that describe the practice in use; a listing of implementation requirements; and an assessment of practice characteristics.

Serves as a systematic means of defining selected practices. Explains the need for and the parts of such a practice profile, and suggests a procedure for the collection of data and development of an accurate profile.

Cost: $3.00

Contact Information:
Cynthia Connolly
The NETWORK, Inc.
290 South Main Street
Andover, MA 01810
(617) 470-1080

Title: An Action Guide to School Improvement

Author(s): Susan Loucks-Horsely and Leslie F. Hergert

Description & Purpose: Intended primarily for educators who are responsible for local school improvement (principals, coordinators, teachers, department heads, administrators); also for trainers, outside support (such as consultants and agencies), and decision makers (e.g., local school boards. education agencies, and legislators).

Describes a practical, research-based conceptual framework for implementing change to bring about school improvement.

The guide is divided into seven linear steps:

1. Establishing the School Improvement Project
2. Assessment and Goal Setting
3. Identifying an Ideal Solution
4. Preparing for Implementation
5. Implementing
6. Review
7. Maintenance and Institutionalization

An appendix is provided, with sources of school improvement resources that are available to schools.

Cost: $5.00

Contact Information:
Cynthia Connolly
The Regional Laboratory for Educational Improvement of the Northeast and Islands
290 South Main Street
Andover, MA 01810
(617) 470-0098

Title: Achieving Excellence (A+)

Sponsor: Mid-Continent Regional Educational Laboratory

Description & Purpose: The Achieving Excellence (A+) Program provides a management tool that helps educators take stock of their current practices in light of current research. A+ suggests performance indicators that local personnel use to gather baseline data. These indicators are then used to monitor the impact of improvement strategies.

Materials: Three ring binder Achieving Excellence, divided into three main sections: Academic Efficiency, Student Success, Improvement Management.

Cost: Dependent upon number of participants and design of program, on a district-by district basis; binder can be purchased separately for $50.00.

Contact Information:
Toni Hass
12500 East Iliff Avenue
Suite 201
Aurora, CO 80014
(303) 337-0990

Titles:

How to Design a Program Evaluation
How to Measure Achievement
How to Measure Program Implementation
How to Present An Evaluation Report

Author(s): Lynn Lyons Morris & Carol Taylor Fitz-Gibbon

Description & Purpose:

Guides and assists practitioners' (at all levels in planning and managing of experience) evaluations.

Consists of four selections from the Program Evaluation Kit (Sage Publications, Beverly Hills). The complete Program Evaluation kit contains eight books that answer questions that a practicing evaluator might ask.

These four practical guides have been field tested and offer details advice, clear definitions, and useful procedures, in non-technical language.

Cost:

Can be purchased as a set ($59.95) or individually:

How to Design a Program Evaluation - $8.50

How to Measure Achievement - $8.50

How to Measure Program Implementation - $7.95

How to Present An Evaluation Report - $4.95

Contact Information:

Sage Publications
275 South Beverly Drive
Beverly Hills, CA 90212
(212) 274-8003

Title: Research Summaries and a Director of Programs, Services, and Resources

Author(s): Lawrence W. Lezotte and Stuart C. Rankin

Description & Purpose: This document is the outcome of the project titled, "Research Dissemination Through Collaborative Planning for School Improvement" (NIE). The first section of the book contains 36 summaries of research helpful to anyone involved in the school improvement planning process for use on any level by educators within the school system, and by consultants, trainers, or outside evaluators. Sections include: Administrative-Instructional Leadership; Monitoring Pupil Progress; School Climate; School Goals and Resources/Basic School Skills; Teacher Expectations; Classroom Management; Instructional Strategies; and Home-School Relations.

The second section contains six case studies that track the project as it occurred in each of the six schools (three elementary level, two middle schools and one high school). Provides descriptions, and contains lessons helpful at any stage of the school improvement process, from planning to implementation to evaluation.

The final section contains resources (local) that illustrate how individual schools might develop a directory of available program resources.

Materials: Case studies, evaluation checklists, planning outlines, Directory of Programs, Services, and Resources, School Improvement Bibliography, Publications List (The Institute for Research on Teaching).

Cost: $15.00 (check made payable to Michigan State University)

Contact Information:
Lawrence W. Lezotte or Stuart C. Rankin
Effective Schools Office
403 Erickson
College of Education
Michigan State University
East Lansing, MI
(571) 353-6413

Title: Developing and Institutionalizing a Self-Improvement Capability: Structures and Strategies of Secondary Schools

Author(s): Herbert J. Klausmeier

Description & Purpose: Developing and Institutionalizing a Self-Improvement Capability (September 1982 - November 1984 study) primarily clarifies the institutionalization of a school's self-improvement capability and also validates the improvement strategies, facilitative organizational structures, and support arrangements involved in a school's initial development of a self-improvement capability. Includes 10 schools (elementary, middle, high school) who annually collected student performance data for 1981-82 and then 1983-84 (academic achievement in one or more areas as measured by standardized tests, average daily attendance, incidence of discipline referrals and suspensions). Reporting forms, observations, checklists, and interviews were used to gather data. Forms are available from the Wisconsin Center for Educational Research. Eight organizational structures and support arrangements were validated as facilitating the implementation of the three improvement strategies: 1) organization of principal and representative teachers into an effectively functioning leadership-coordinating group. This is most critical because it is essential for implementing the goal setting strategy and is prerequisite to effective functioning of the remaining seven; 2) organization of teachers and students into small groups; 3) teachers advising students; 4) scheduling of classes so that mutual interests have a common time during the school day for planning improvement activities; 5) principal's effective instructional leadership; 6) ongoing inservice/staff development; 7)support of district officials; 8) school's autonomy to implement improvement program; and 9) parent participation and support.

Cost: $12.25

Contact Information:
University Press of America
4720 Boston Way, Lanham, MD 20706
(301) 359-3366

Additional Books

Education USA: Special report, Good Schools: What Makes Them Work (Virginia: National School Public Relations Association), 1981.

Brief book that provides clear and concise overview of the literature, philosophy, and procedures of effective schools movement.

Edgar A. Kelly, Improving School Climate: Leadership Techniques for Educators. (National Association of Secondary School Principals), 1980.

The central theme of this monograph is that school can make a difference in what happens to the people who work and study in them. Focusing on the role of the principal in affecting school climate, it deals most directly with assessing school climate and planing for its improvement.

Rolf Lehming and Michael Kane (ed.), Improving Schools: Using What We Know (New York: Sage Publications), 1981.

Series of articles written by such noted people as Ernest House and Matthew Miles, the work was developed under a contract with the Far West Laboratory and funding was provided by NIE. The articles are concerned with innovation and organizational change in the schools.

George F. Madau, Peter W. Airasian, and Thomas Kellaghan, School Effectiveness: A Reassessment of the Evidence (New York: McGraw-Hill), 1980.

This book discusses some of the major issues in determining what about schools really makes a difference in student learning. Some of the key literature in local school improvement is summarized.

Carl L. Marburger, One School at a Time (Columbia, MD: The National Committee for Citizens in Education), 1985.

Written by one of the key figures in the movement, this book is designed to familiarize various members of the school community with the ideas and concepts of school-based management and help them to get started with its implementation.

Daniel C. Neale, William J. Bailey, and Billy E. Ross, Strategies for School Improvement: Cooperative Planning and Organization Development (Boston: Allyn and Bacon), 1981.

This book "written especially for educators who are seeking a careful summary of current knowledge about planned change in education," focuses on the local school-building organization and collaboration among all members of the local school community as key ingredients for change. Using the "Partnership Model for School Improvement," topics such as the tactics and methods of

school improvement are discussed. Bibliographies, planning guides, and questions for discussion are included.

Manuals

California State Department of Education, Assistance Guide for Forming a Consortium to Improve School Programs (Sacramento, CA: State Department of Education), 1981.

Manual for districts covers all facets of consortium formation.

National Study of School Evaluation, Evaluation Criteria: Fifth Edition (Virginia: NSSE), 1978.

The manual is based "on the principle that a school should be judged in terms of what it is striving to achieve and according to the extent to which it is meeting the needs of the students enrolled and the community it serves." It provided instructions, procedures, and sample forms for self-evaluation.

Seattle Public Schools, School-Based Planning Manual, 1982.

A step-by-step planning guide including forms and checklists.

Organizational Resources - Regional Educational Laboratories, R&D Centers, Universities

The Academy for Educational Development, School Services Division, 680 Fifth Avenue, New York, NY 10019, Telephone (212) 397-0040.

Technical assistance and resource materials are available.

Iowa State University, School Improvement Model, College of Education, Ames, IA 50011, Telephone: (515) 294-5521

A method of school-based assessment developed from the effective schools literature. Consulting services may be contracted. Extensive manuals and related literature are available.

National Study of School Evaluation, 5201 Leesburg Pike, Falls Church, VA 22041.

Can provide such resources as evaluation manuals, and inventories of parent, teacher, and student opinions.

Northwest Regional Educational Laboratory, 300 SW Sixth Ave., Portland, OR 97204, Telephone: (800) 547-6339.

Onward to Excellence Program was developed to train leadership teams in schools to apply effective schools research and goal based management practices to local improvement efforts. Program descriptions, manuals, and resource lists are available.

Research for Better Schools (RBS), 444 North Third St., Philadelphia, PA 19123, Telephone: (215) 574-9300.

The Secondary School Development Program and School Effectiveness Training Program aim to improve effectiveness at the school building level. RBS provides technical assistance, materials, manuals, and survey instruments. These programs are jointly sponsored by the New Jersey Education Association and Pennsylvania State Education Association.

The Regional Laboratory for Educational Improvement of the Northeast and Islands, 290 South Main Street, Andover, MA 01810, Telephone: (617) 470-0098.

Technical assistance and resource materials are available.

Organization Resources - School Districts and State Departments of Education

California State Department of Education, Office of School Improvement, 721 Capitol Mall, Sacramento, CA 95814, Telephone: (916) 322-5954.

One of the oldest and most extensive statewide efforts. Available materials include handbooks, manuals, and evaluation reports.

Chicago Public Schools, Chicago Effective Schools Project, 1819 W. Pershing Road, East Center 6, Chicago, IL 60609, Telephone: (312) 890-7771.

In operation since 1981, the program strives to apply effective schools research in instituting a program to lessen the educational disadvantage that accrues to minority students in racially isolated schools. They can provide technical assistance, plans, and evaluation reports.

New York City Public Schools, Office of Comprehensive School Improvement Projects (OCSIP), 110 Livingston Street, Brooklyn, NY, Telephone: (718) 935-4055.

Based on the principles of school effectiveness as articulated by such researchers as Ron Edmonds, the program has been implemented in elementary and middle schools throughout the city. Technical assistance and instruments are available. In addition evaluation reports and documentation instruments are available through the Office of Educational Assessment, 110 Livingston Street, 7th Floor, Brooklyn, NY 11201.

Appendix 4

Bibliography of School Effectiveness Studies

Alexander, W. M. & George, P. S. (1981). *The exemplary middle school*. New York: Holt, Rinehart, and Winston.

Averch, H. A., Carroll, S. J., Donaldson, T. S., Kiesling, H. J., & Pincus, J. (1972). *How effective is school? A critical review and synthesis of research findings*. Santa Monica, CA: Rand Corporation.

Azumi, J. and Madhere, S. (1982). *Characteristics of high achieving elementary schools in Newark*. Newark School District Report, ORET Report No. 5. Newark, NJ: Office of Research, Evaluation and Testing, Newark Board of Education.

Barker, R. G., & Gump, P. V. (1964). *Big school, small school*: *High school size and student behavior*. Stanford, CA: Stanford University Press.

Bell,D. (1972). On meritocracy and quality. *Public Interest*, *29*, 29-68.

Bereiter, C. (1963). Some persisting dilemmas in the measurement of change. In *Problems in Measuring Change*. Madison: University of Wisconsin Press.

Bereiter, C. (1969). The future of individual differences. *Harvard Educational Review*, *39*, 310-318.

Bidwell, C. E., and Kasarda, J. D. (1975). School district organization and student achievement. *American Sociological Review*, *40*, 55-70.

Bidwell, C. E., and Kasarda, J. D. (1977). Conceptualizing and measuring the effects of schools and schooling.

Bidwell, C. E., and Kasarda, J. D. (1980). Conceptualizing and measuring the effects of school and schooling. *American Journal of Education*, *88*, 401-430.

Blalock, H. M. Jr. (1974). Beyond ordinal measurement: Weak tests of stronger theories. *Measurement in the Social Sciences*. Chicago: Aldine.

Block, A. W. (1983). *Effective schools*: *a summary of research*. Arlington, VA: Educational Research Service.

Bloom, B. S. (1963). Testing cognitive ability and achievement. *Handbook of research on teaching*. Chicago: Rand McNally.

Blumbert, A., and Greenfield, W. (1980). *The effective principal: perspectives on school leadership*. Boston: Allyn and Bacon.

Bowles, S., and Levin, H. M. (1968). More on multicollinearity and effectiveness of schools. *Journal of Human Resources*, *3*, 393-400.

Boyer, E. (1983). *High school: a report on secondary education in America*. New York: Harper & Row.

Bridge, R. G., Judd, C. M., and Moock, P. R. (1979). *The determinants of educational outcomes: the impact of families, peers, teachers, and schools*. Cambridge, MA: Ballinger.

Brookover, W. B. (1981). *Effective secondary schools*. Philadelphia, PA: Research for Better Schools.

Brookover, W. B., Beady, C., Flood, P., Schweitzer, J., and Wisenbaker, J. (1979). *School social systems and student achievement: schools can make difference*. New York: Praeger.

Brookover, W. B., Beady, C., Flood, P., Schweitzer, J. & Wisenbaker, J. (1979). *School social systems and student achievement*. New York: Praeger.

Brundage, D. (Ed.). (1980). *The journalism research fellows report: what makes an effective school*? Studies of schools that work in: Arkansas, Florida, Maine, Maryland, Nebraska, Virginia, Urban Districts Nationwide. Washington, D.C.: Institute for Educational Leadership, George Washington University.

California State Department of Education. (1977). *California school effectiveness study, the first year: 1974-74*. Sacramento, CA: Office of Program Evaluation and Research.

Clark, D. L., Lotto, L. S., and McCarthy, M. M. (1980). Factors associated with success in urban elementary schools. *Phi Delta Kappan*, *61*, 467-470.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J. M. Mood, A., Weinfield, F. D., & York, R. L. (1966). *Equality of Educational Opportunity*. Washington, D.C.: U. S. Government Printing Office.

Coleman, J. S. (1968). The concept of equality of educational opportunity. *Harvard Educational Review*, *38*, 7-22.

Coleman, J. S., Hoffer, T., & Kilgore, S. (1982). *High school achievement: public, Catholic, private schools compared*. New York: Basic Schools.

Corcoran, T. B., and Wilson, B. L. (ND). *Characteristics of exemplary high schools*. Philadelphia, PA: Research for Better Schools.

Crain, R. L., & Ferrer, R. L. (1982). Achievement prediction with school-level equations: a non-technical example using the public and private school data (Report 323). Baltimore: John Hopkins University, Center for Social Organization of Schools.

Cronbach, L. J., and Furby, L. (1970). How should we measure 'change' - or should we? Psychological Bulletin, 74, 68-80.

Cuban, L. (1984). Transforming the frog into a prince: Effective schools research, policy and practices at the district level. Harvard Educational Review, 54(2), 129-151.

Cuervo, A. G. (Ed.). (1984). Toward better and safer schools. Alexandria, VA: National School Boards Association.

Cusick, P. (1983). The egalitarian ideal and the American high school. New York: Longman.

D'Amico, J. (1982). Using effective school studies to create effective schools: No recipe yet. Educational Leadership, 40(3), 60-62.

Ebert, R. W., Kehoe, E., & Stone, J. A. (1984). The effect of school size on student outcomes. Eugene, OR: University of Oregon, Center for Educational Policy and Management (ED 245 382).

Edmonds, R. (1979). Some schools work and more can. Social Policy, 9, 32.

Edmonds, R. (1979). Effective schools for the urban poor. Education Leadership, 37, 1.

Edmonds, R. (1981). Making public schools effective. Social Policy, 12, 56-60.

Edmonds, R. (1982). Programs of school improvement. Educational Leadership, 40, 3.

Edmonds, R. R., and Frederiksen, J. R. (1979). Search for effective schools: the identification and analysis of city schools that are instructionally effective for poor children. Cambridge, MA: Center for Urban Studies, Harvard University.

Educational Research Service. (1977). Student absenteeism. ERS Report. Arlington, VA: Author.

Educational Research Service. (1983). Organization of the middle grades: A summary of research. Arlington, VA: Author.

Eubanks, E. E., and Levine, D. U. (1983). A first look at effective school projects in New York City and Milwaukee. Educational Leadership, 64, 10.

Farrar, E., Neufeld, B., & Miles, M. (1984). Effective school programs in high schools: Social promotion or movement by merit? Phi Delta Kappan, 65, 701-706.

Fetter, W. B., Collins, E. F., & Smith, J. W. (1968). Characteristics differentiating under- and overachieving elementary schools. Technical Note No. 63. Washington, D.C.: Division of Data Analysis and Dissemination, National Center for Education Statistics.

Finn, C. E. (1984). Towards strategic independence: Nine commandments for enhancing school effectiveness. Phi Delta Kappan, 65, 8.

Firestone, W. A., and Herriott, R. E. (1982). Prescriptions for effective elementary schools don't fit secondary schools. Educational Leadership, 40(3), 51-53.

Firestone, W. A. and Wilson, B. L. (1983). Using bureaucratic and cultural linkages to improve instruction: the high school principal's contributions. Philadelphia, PA: Research for Better Schools.

Ford Foundation. (1984). City high schools: a recognition of progress. New York: Ford Foundation.

Glasman, N. S. and Biniaminov, I. (1981). Input-output analyses of schools. Review of Educational Research, 51, 509-539.

Good, T. L. (1983). Classroom research: A decade of progress. Educational Psychologist, 18(3), 127-144.

Goodlad, J. I. (1983). A place called school. New York: McGraw-Hill.

Gottfredson, G. D., & Gottfredson, D. C. (in press). Victimization in six hundred schools: an analysis of the roots of disorder. New York: Plenum.

Grabe, M. (1981). School size and the importance of school activities. Adolescence, 16(61), 21-31.

Guthrie, J. W. (1970). A survey of school effectiveness studies, in Do teachers make a difference?, pp. 25-54. Washington, D.C.: Bureau of Educational Personnel Development, Office of Education, U. S. Department of Health, Education and Welfare.

Heise, D. R. (1969). Separating reliability and stability in test-retest correlation. American Sociological Review, 34, 93-101.

Herriott, R. E., & Muse, D. N. (1973). Methodological issues in the study of school effects. In F. N. Kerlinger (Ed.), Review of Research in Education: Vol. 1 (pp. 209-36). Itasca, IL: Peacock.

Herzog, Elizabeth, Newcomb, Carol H., and Cisin, I. H. (1974). Double deprivation: The less they have, the less they learn. *A Report on Longitudinal Evaluations of Preschool Programs*. Washington, D. C.: Department of Health, Education and Welfare.

Heyns, B. (1976). Education, evaluation, and the metrics of learning. *Journal for Teaching and Learning*, *2*, 2-16.

Heyns, B. (1976). Exposure and the effects of schooling. *Final Report, Grant No. 10301*. Washington, D.C.: National Institute of Education.

Heyns, B. (1977). Measuring the effects of education. *Science*, *196*, 763-765.

Heyns, B. (1978). *Summer learning and the effects of schooling*. New York: Academic Press.

Hodgson, G. (1975). Do schools make a difference? *The Inequality Controversy*. New York: Basic Books.

Hopkins, K. (1982). The unit of analysis: Group means versus individual observations. *American Educational Research Journal*, *19*, 5-18.

Hyman, H. H., Wright, C. R., and Reed, J. S. (1975). *The enduring effects of education*. Chicago, IL: University of Chicago Press.

Jencks, C., &Brown, J. (1975). The effects of high schools on their students. *Harvard Educational Review*, *45*, 273-324.

Jencks, C., Smith, M., Ackland, H., Bane, M., Cohen, D., Gintis, H., Heyns, B., & Michelson, S. (1972). *Inequality: a reassessment of the effect of family and schooling in America*. New York: Harper and Row.

Karweit, N. (1985). Should we lengthen the school term? *Educational Researcher*, *14*, (6), 9-15.

Kimbrough, J. & Hill, P. T. (1981). *The aggregate effects of federal education programs*. Santa Monica, CA: Rand Corp.

Kirst, M. W. (1983). Effective schools: Political environment and education policy. *Planning and Change*, 234-244.

Klitgaard, R. E., and Hall, G. (1973). *Are there unusually effective schools*? Santa Monica, CA: Rand.

Kulik, C. & Kulik, J. (1982). Effects of ability grouping on secondary school students: A meta-analysis of evaluation findings. *American Educational Research Journal*, *19*, 415-428.

Lezotte, L. W. (1982). A response to D'Amico: Not a recipe but a framework. *Educational Leadership*, *40*, 63.

Lightfoot, S. L. (1983). The good high school: portraits of character and culture. New York: Basic Books.

Lindsay, P. (1982). The effect of high school size on student participation, satisfaction and attendance. Educational Evaluation and Policy Analysis, 4(1), 57-66.

Lipsitz, J. (1984). Successful schools for young adolescents. New Brunswick, NJ: Transaction Books.

McDill, E. L., Natriello, G. & Pallas, A. M. (1985). Raising standards and retaining students: the impact of the reform recommendations on potential dropouts. (Report No. 358). Baltimore: John Hopkins University, Center for Social Organization of Schools.

Miles, M. B. (1981). Mapping the common properties of schools. In R. Lehming & M. Kane (Eds.), Improving schools: using what we know (pp. 42-114). Beverly Hills, CA: Sage.

Morgan, W. R. (1983). Learning and student life quality of public and private schools youth. Sociology of Education, 56(4), 187-202.

Murnane, R. J. (1975). The impact of school resources on the learning of inner city children. Cambridge, MA: Ballinger.

Murnane, R. J. (1980). Interpreting the evidence on school effectiveness. Working paper No. 830. New Haven, CT: Institution for Social and Policy Studies, Yale University.

Murphy, J. F. et al. (1982). Academic Press: Translating high expectations into school policies and classroom practices. Educational Leadership, 40, 22-26.

National Commission on Excellence in Education. (1983). A nation at risk: the imperative for educational reform (A report to the Nation and the Secretary of Education). Washington, D.C.: U. S. Department of Education. (ERIC No. ED 226 006).

National Education Association. (1968). Ability grouping. Research Summary 1968-83. Washington, D.C.: NEA Research Division.

Neufeld, B., Farrar, E., and Miles, M. B. (1983). A review of effective schools research: the messages for secondary schools. Prepared for the National Commission on Excellence in Education. Cambridge, MA: The Huron Institute.

Phi Delta Kappan. (1980). Why do some urban schools succeed? Bloomington, IN: Phi Delta Kappan.

Purkey, S. C., & Smith, M. S. (1985). School reform: The district policy implications of the effective schools literature. Elementary School Journal, 85, 353-389.

Purkey, S. C., & Smith, M. S. (1983). Effective schools: A review. *Elementary School Journal*, *83*, 427-452.

Ralph, J. H. & Fennessey, J. (1983). Science or reform: Some questions about the effective schools model. *Phi Delta Kappan*, *64*, 689-694.

Rock, D. A., Ekstrom, R. B., Goertz, M. E., & Pollock, J. M. (1985). *High school excellence*. Princeton, NJ: ETS.

Rosenholtz, S. J. (1985). Effective schools: Interpreting the evidence. *American Journal of Education*, *93*, 352-388.

Rowan, B., Bossert, S. T., & Dwyer, D. C. (1983). Research on effective schools: A cautionary note. *Educational Researcher*, *12*, (April), 24-31.

Rutter, M., Maughan, B., Mortimore, P., Ouston, J., & Smith, A. (1979). *Fifteen thousand hours: secondary schools and their effects on children*. Cambridge, MA: Harvard University Press.

Simmons, R. G., et. al. (1979). Entry into early adolescence: The impact of school structure, puberty, and early dating on self-esteem. *American Sociological Review*, *44*, 948-967.

Sizer, T. (1984). *Horace's compromise: the dilemma of the American high school*. Boston: Houghton-Mifflin.

Slavin, R. E. (1984). Meta-analysis in education: How has it been used? *Educational Researcher*, *13*, 6-15,m 24-27.

Sorensen, A. B., & Hallinan, M. T. (1977). A reconceptualization of school effects. *Sociology of Education*, *50*, 273-289.

Stiegelbauer, S. M. (1984). *Community, context and co-curriculum: situational factors influencing school improvements in the study of high schools*. (R & D Report No. 3186). Austin, TX: Research and Development Center for Teacher Education.

Summers, A. A., & Wolfe, B. L. (1977). Do schools make a difference? *American Economic Review*, *65*, 639-52.

Summers, A. A., & Wolfe, B. L. (1975). Which school resources help learning? Efficiency and equity in Philadelphia public schools. Federal Reserve Bank of Philadelphia. *Business Review*.

Toch, T. (1984). The dark side of the excellence movement. *Phi Delta Kappan*, *66*, 173-176.

Twentieth Century Fund. (1983). *Making the grade*. New York: Author.

Tyack, D. B., Kirst, M. W., and Hansot, E. (1980). Educational reform: Retrospect and prospect. Teachers' College Record, 81(3), 253-269.

U. S. Department of Health, Education and Welfare. (1978). Violent Schools - safe schools. Vol. 1. Washington, D.C.: Author.

Weber, G. (1971). Inner City children can be taught to read: four successful schools. Washington, D.C.: Counsel for Basic Education.

Wellisch, J. B., et. al. (1978). School management and organization in successful schools. Sociology of Education, 51, 211-226.

Willms, J. D. (1983). Do private schools produce higher levels of academic achievement? New evidence for the tuition tax credit debate. In T. James and H. M. Levin (Eds.), Public dollars for private schools: the case of tuition tax credits. Philadelphia: Temple University Press.

Wynne, E. A. (1980). Looking at schools: Good, Bad, and Indifferent. Lexington, MA: Heath.